AGENT-BASED MODELLING AND SIMULATION OF COLLABORATIVE CONTENT CREATION: THE CURIOUS CASE OF WIKIPEDIA

ERŞAN TAŞAN

BOGAZİÇİ UNIVERSITY

2014

AGENT-BASED MODELLING AND SIMULATION OF COLLABORATIVE CONTENT CREATION: THE CURIOUS CASE OF WIKIPEDIA

Thesis submitted to the
Institute for Graduate Sciences in the Social Sciences
in partial fulfilment of the requirements for the degree of

Master of Arts
in
Management Information Systems

by

Erşan Taşan

Boğaziçi University

2014

Abstract

Erşan Taşan, "Agent-Based Modelling and Simulation of Collaborative Content Creation: The Curious Case of Wikipedia"

Wikipedia is no doubt one of the most important collaborative informational products of human history. The collaborative effort that produces Wikipedia is of fundamental importance because the community includes anonymous, uncompensated editors and a lack of observable top-down hierarchy.

In this study, we propose an agent-based model for Wikipedia users' activity of collaborative article editing. In this graph-theoretical approach every user and article are represented as vertices in a multimodal affiliation network. When a user chooses to edit an article, an edge between the node of the user in question and that of the edited article, is created. User preferences, statuses, relative content quality of articles, distribution of collaboration and resulting relationships are examined in the network.

We analyse input parameters' effects on resulting principal graph characteristics, namely the clustering coefficient, the path length, the small world characteristic Q, degree correlation, and degree distribution. Simulation findings point out that, users' area of interest dimensions and active user percentage are positively correlated with the total edge count in the graph; therefore, the encyclopaedia quality. Conversely, good article threshold parameters raise high-quality article specifications and are negatively correlated with the total edit count in the encyclopaedia.

We recommend an easier, automated process for the selection of good and featured articles of Wikipedia. Experiments have demonstrated that lowering the barrier of high quality status for articles, results in more effort and quality for the encyclopaedia as a whole. Additionally, we recommend more internal link concentration in good and featured articles, in order to spread the effort of their successful editors.

Tez Özeti

Erşan Taşan, “Agent-Based Modelling and Simulation of Collaborative Content Creation: The Curious Case of Wikipedia”

Vikipedi (Wikipedia), şüphesiz ki insanlık tarihinin en önemli müşterek bilgi ürünlerinden biridir. Vikipedi’yi üreten ortak çaba temel bir önemi haizdir çünkü, topluluk anonim, ücret almayan editörleri içermekte ve gözlenebilir bir yukarıdan aşağıya hiyerarşi bulunmamaktadır.

Bu çalışmada biz, Vikipedi kullanıcılarının ortaklaşa metin/madde düzenleme faaliyetinin bir ajan-bazlı modelini öneriyoruz. Bu şebeke (graph) teorisi yaklaşımında, her kullanıcı ve madde çokşekilli (multimodal) ilişki ağında birer düğüm (vertex) ile temsil edilmektedir. Eğer bir kullanıcı bir maddeyi düzenlemeyi seçerse, o kullanıcının düğümü ile ilgili maddenin düğümü arasında bir bağ (edge) yaratılır. Kullanıcı tercihleri, statüleri, maddelerin nispi içerik kaliteleri, yardımlaşmanın dağılımı ve sonuçta oluşan ilişkiler, ağ üzerinde incelenir.

Girdi parametrelerinin, topaklanma katsayısı (clustering coefficient), yol uzunluğu (path length), küçük dünya karakteristiği (small world characteristic), derece orantısı (degree correlation) ve derece dağılımı (degree distribution) gibi başlıca şebeke niteliklerine etkisi analiz edilmiştir. Benzetim (simülasyon) bulguları, kullanıcıların ilgi alanlarının boyutlarının ve aktif kullanıcı yüzdesinin, şebekedeki toplam bağ sayısı ile, dolayısıyla ansiklopedi kalitesiyle, doğru orantılı olduğuna işaret etmektedir. İyi madde eşik parametreleri ise, aksine, yüksek kaliteli içerik (iyi madde ve seçkin madde) eşiğini yükseltmekte ve ansiklopedideki toplam düzenleme sayısı ile ters orantılı olduğu görülmektedir.

Vikipedi’de iyi ve seçkin madde seçimi için kolaylaştırılmış ve otomatik bir süreç tavsiye ediyoruz. Deneyler seçkin madde eşiğinin düşürülmesinin Vikipedi’nin tamamında kalite artışı sağladığını göstermektedir. Ayrıca başarılı editörlerin emeğinin yayılması için seçkin içerikte içsel bağlantı yoğunluğunun arttırılmasını tavsiye etmekteyiz.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In this study, we mainly target at understanding and explanation of inner structure and self-organisation of Wikipedia, utilising agent-based modelling and simulation (ABMS) methodology, which appeared as a need, and has been encouraged previously in the literature (Ingawale, 2008). Additionally, we measure effects of several agent-based variables; for instance aspects of Wikipedia community members such as having diverse interests, or regulations by Wikipedia policy-makers such as good/featured article criteria; on the resulting encyclopaedia quality and structure. Our agent-based Wikipedia model is based on graph theory and the quality and structure of resulting network are determined by well-defined graph theory metrics such as clustering coefficient or small-world characteristic Q.

We introduce agent-based modelling and simulation (ABMS) and Wikipedia concepts and explain their history briefly first. Then we explain previous efforts for modelling and simulating Wikipedia and why such a model is useful in diverse areas. Agent-based modelling and graph theory software are described and Repast Symphony application suite is introduced. In model section, we examine Anthony et al.'s (2005) influential grouping for Wikipedia community, namely *Good Samaritans* (anonymous one-shot contributors) and *Zealots* (repetitively contributing members of online community that seek reputation). We further categorise zealot users into two classes, according to their career preferences: *Project Leaders* (specialists who make regular contributions on their topic of expertise) and *Administrators* (active users who tidy up articles and ensure they abide by the article guidelines). We discuss how to apply these community groupings in our model on a graph and introduce Cartesian Plane of Information notion. Because articles and users are represented by nodes in the graph, we implement the edge concept as the cooperation and relationship symbol. We build edges, when a user edits an article, between user and article, between user and previous editors of the article, and between this article and other articles edited by the user (in three separate graphs). We then define parameters for this model, as distribution of these three basic classes in community, diversity of interest of users for editing, and selection criteria for an article to be defined as "Good" (featured).

The system dynamics that construct Wikipedia are of particular interest because; it resembles no other system we have seen before. A considerable number of edits are made by anonymous users ("User: Opabinia regalis/Article statistics", 2007), there is no theoretical top-down hierarchy in the community ("Wikipedia: About", n.d.), and even the most valuable articles of the encyclopaedia might be deleted by a vandal user in one step. And this mechanism of collaboration produces the most consulted multilingual encyclopaedia ever created. Additionally, this novel collaboration technique might be utilised in numerous real-life situations from knowledge management systems (KMS) in companies to prepare training documents for new employees by existing seasoned employees, to constitutions being prepared by the governments, in order to include the nation in the process.

As we attempted to understand and comprehend the driving forces behind Wikipedia, we observed that the completely novel and valuable phenomenon is not the product itself (an encyclopaedia), but the human behaviour that created it. Human species has invented a new mechanism to collaborate and cooperate anonymously all over the world. It is clear that the free encyclopaedia concept comprehension and explanation efforts should focus on this novel cooperation method among human beings. They should focus the human beings who cooperatively build Wikipedia are its users, the community members, the agents of the system. Therefore in our simulation efforts, our model for Wikipedia collaboration process is agent-based.

Agent-based modelling and simulation (ABMS) is a relatively recent approach in systems modelling and simulation, in which the observer models and simulates actions and interactions of *agents* (meaningful autonomous units of the system), in order to deduce and determine the structure of the whole system (Izquierdo et al., 2008). Individual agents may act on a defined set of rules or according to what they perceive as their own interests, such as reproduction, economic benefit, or social status, with bounded rationality or limited information. They may experience learning, adaptation and reproduction as well (Macal & North, 2008). This class of computational modelling, in which agents are main actors, has been of service in providing insights about many diverse areas from military doctrines (Cioppa, Lucas & Sanchez, 2004) to wildfires (Hernández Encinas, 2007); from optimization of production (Holmgren, Persson & Davidsson, 2013) to evacuation model of supermarket (Shuangyun, Kun, Quanli & Yuhua, 2010).

Wikipedia is a collaboratively created, community-led free encyclopaedia that anyone can edit. It is referred as one of the most prominent products of cooperative effort in human history (Yasseri, 2012), and the success of altruistic collaboration (Kuznetsov, 2006) of supposedly democratic-anarchic nature (Muller-Seitz & Reger, 2010). A 2005 Nature study found Wikipedia's accuracy rate as nearly equal to that of Encyclopædia Britannica (Giles, 2005), whereas a more comprehensive study measured better quality and accuracy (Casebourne et al., 2012)

Wikipedia stems from “wiki” concept, which means user-editable web page. It is a relatively new concept first introduced by Cunningham as the wikiwikiweb project (1995). Wikipedia was born in 2001, and in less than fifteen years, it became a multi-language online encyclopaedia that covers 24.6 million articles in 285 languages (Wikipedia: Statistics, 2013).

Regarding agent (users) oriented point of view for Wikipedia, and modelling it to understand it, there seems to be a recent scholarly interest in the literature. In their influential paper Anthony et al. (2005) classify Wikipedia community members into two main groups: Good Samaritans and Zealots. There are studies proposing to reach the fusion between these academically popular entities, Wikipedia and agent-based modelling discipline, in order to shed the light into almost-flawlessly working mechanism of the free encyclopaedia. Ingawale (2008) exactly targets, and calls for the subject matter of our study in his “Understanding the Wikipedia Phenomenon: A Case for Agent Based Modelling” article and argues that such a model, if applied, may lead to useful insights about self-organisation in Wikipedia.

Because Wikipedia is an output of autonomous and asynchronous-behaving agents (encyclopaedia editors and users) interacting with each other, it is a good example of emergence. It is difficult to predict that collaboration of non-professional internet-interested individuals might produce a massive encyclopaedia and a social phenomenon. With its community centred self-organization and self-healing capabilities, Wikipedia is a well organized representation of its users’ and contributors’ knowledge, thus, modelling it based on these users and contributions will approach its true nature better, bringing about more meaningful results (Ingawale, 2008). From organizations training new employees to universities creating textbooks, to even states preparing a constitution, if entities find the ways of harnessing the power of the crowd, this would result in a more powerful content creation. This may enable better organization in Wikipedia itself as well – for instance we might point out which areas needs editing or which projects lack adequate contributors.

CHAPTER 2

BACKGROUND

In this chapter, to introduce the paradigm of Agent-Based Modelling and Simulation (ABMS), background information on basic ABMS concepts and roots is presented first. Then Wikipedia studies and collaborative content creation subjects (specifically in virtual and online environments) are summarised. Lastly, the modelling efforts for Wikipedia, the use of ABMS in simulation of collaborative content creation and associated literature review is introduced.

Agent-Based Modelling and Simulation

Agent-Based Modelling and Simulation (ABMS) is a class of computational models that allows the simulation of actions and interactions of autonomous agents (or individuals) in an environment, in order to determine what effects occur in the whole system (Izquierdo et al., 2008).

ABMS is closely linked to, and combines elements of games theory, complex systems, emergence, computational sociology, multi agent systems, evolutionary programming and cellular automata (Solomon, 2013). The models simulate the simultaneous or sequential operations and interactions of multiple entities (agents), in an attempt to recreate and predict the appearance of complex phenomena. It is a process of emergence from micro to macro levels. Individual agents act on a defined set of rules or according to what they perceive as their own interests, such as reproduction, economic benefit, or social status, with bounded rationality or limited information. The ABMS agents' sets of rules need not be fixed and agents may experience learning, adaptation and reproduction (Macal & North, 2008).

ABMS technique has illuminated many major phenomena and led to new questions in many diverse areas from military doctrines (Cioppa, Lucas & Sanchez, 2004), to segregation of society (Schelling, 1971); from pedestrian flow and movement (Turner & Penn, 2002), to wildfires (Hernández Encinas, 2007), although in biological and ecological field studies individual-based model (IBM) term is used instead (Grimm & Railsback, 2005). The agent-based model concept has been developed to explore complex systems in which an entire system is greater than sum of its parts (Lewin, 1992; Holland, 1995). Using such models, in which the system is viewed as an organism consisting of smaller independent units, the emergent behaviour of complex networks such as the Internet, neurons communicating together to create the miraculous human brain (Smith, 1996; 1998) and sudden developments in the stock market are analysed (Giardina & Bouchaud, 2003). Agent-based models have been utilised since the mid-nineties as a practical problem solution method as well. They have been applied to diverse business or technological problems from optimise production (Holmgren, Persson & Davidsson, 2013), to better understand consumer behaviour (e.g. response to word-of-mouth (Delre, Jager, Bijmolt & Janssen, 2007); evacuation model of supermarket (Shuangyun, Kun, Quanli & Yuhua, 2010), and to even implementation of new agent-based peer-to-peer systems (Babaoğlu, Meling, & Montresor, 2002).

The agent-based modelling has its roots both in the cellular automata models and Von Neumann machine (McMullin, 2000), as well as in the various areas of artificial intelligence. In comparative analysis, the agent-based modelling can be interpreted as an extension of cellular automata. In cellular automata the smallest unit is the cell of a regular lattice; “it is by definition fixed and may hold discrete states. Agents in ABMS on the other hand are much more flexible and may hold several characteristic features (e.g. height, colour, diameter, health status, speed, pressure, etc.)" (Emrich, Breitenecker, Zauner & Popper, 2008). ABMS is a special case of micro-simulation. Agent-based models are based on the theory of multi-agent systems.

An agent-based model is a multi-agent system with a set of autonomous agents that operate in parallel and communicate with each other. The individual properties of the agents describing their behaviour and interactions are called as basic properties, and properties that appear (emerge) in a more collective (system-wide and higher) level are known as *emergent* properties. Emergence is an important aspect for the study of complex systems, and behaviour of the system emerges from the global behaviour of the agents and their interactions when running the simulation. Classically (British emergentism) used to be related to unexpected, unexplainable or in-deducible results (Broad, 1925), but modern definition of emergence adopts a more scientific approach. Axelrod (1997b, p. 194) states "there are some models ... in which emergent properties can be formally deduced."

Emergence may refer to counterintuitive outcomes stemming from interaction of agents in many cases. Traffic congestion is basically an emergent phenomenon of vehicles' interaction. However, while vehicles are heading north, traffic jam moves south, in the opposite direction (Kesting, Treiber & Helbing, 2008). This is because every driver adjusts its speed to new conditions when approaching the rear of the jam. This lengthens jam's tail, while leading cars increase their speed back to normal levels, shortening the frontal part.

Agent-based modelling is argued to be a third way of carrying out science along with argumentation and formalisation in social sciences (Gilbert & Tema, 2000). This argument holds true more generally as well, ABMS might be accepted as the third way along with standard procedures of inductive and deductive reasoning (Axelrod, 1997). Deduction refers to theory building based on a set of logical assumptions while induction includes finding patterns in, and analysing results from the empirical data. Simulation in general and ABMS in particular, covers elements of both. Starting point is the assumptions set for the model (e.g. rules for agents and environment in ABMS), but rather than directly using these, a simulation based on them, is conducted. This process provides empirical data, albeit generated one rather than having real world information. These data then utilised for inductive reasoning and analysed thoroughly (Tesfatsion & Axelrod, 2013).

This revolutionary image is supported by following result counts from the major academic s2earch engines. Agent-based modelling related terms seem to create a novel paradigm, as shown in Table 1 (data retrieved on 16 March 2013):

Table 1 Article Count from Several Academic Search Engines

Sources \ Keywords	agent based model	agent-based model	agent-based
Google Scholar	~ 3,510,000	~ 309,000	~ 391,000
ScienceDirect	731,716	731,716	1,185,422
BASE	36,060	6,953	10,726

In addition to scholarly interest, Google Trends tool (2013) points out a paradigm shift in Google search frequency beginning from 2004, observable in Figure 1, with a relative recent increase in interest (blue line represents search volume statistics for keyword *agent-based model* and red line represents that of *agent based modelling*).

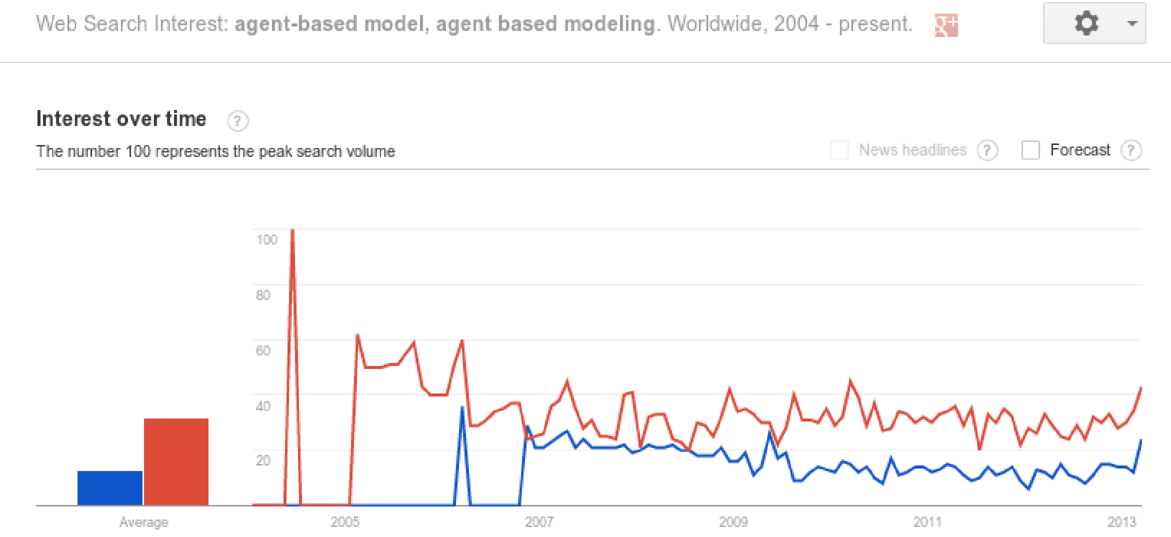


Fig. 1 Web search interest timeline for keywords *agent-based model* (blue line) and *agent based modelling* (red line).

The foundation of ABMS can be traced back to Von Neumann cellular automaton, a conceptual device proposed in 1949, which is able to reproduce itself (self-replicate) (McMullin, 2000). Its original designer Von Neumann prepared the system on a grid (a 2-D lattice), at the suggestion of another mathematician Stanislaw Ulam, generating the first *cellular automata*, a term later coined. John Conway constructed the next major improvement, and introduced the Game of Life (Conway, 1970), with considerably simple rules, in which unpredictable figures (e.g. moving gliders fired by a glider gun) appear on a 2-D lattice. This study is an early example of emergence in a user-defined system, with intuition unable to anticipate the outcome, by just examining the rules.

Another step of fundamental importance in ABMS history is Schelling's segregation model (Schelling, 1971), in which he modelled inhabitant preferences for neighbours of same ethnic origin or colour. Although reiterated with computational power many times later, he prepared original work with pen, paper, nickels and dimes on a physical board. This work opened up a new path in social sciences, in which social experiments are difficult to conduct on mass populations. Rather than experimenting, researchers were now able to model and simulate – a critically acclaimed contribution, which led to Schelling's 2005 Nobel Memorial Prize in Economic Sciences.

Robert Axelrod, whose contributions supplied both breadth and depth to the field, made progress with Dresher and Flood's popular Prisoner's Dilemma Game and hosted a tournament on suggested strategies for long-term gain. (Axelrod, 1980) This historical period of agent-based simulation earned popularity to the field, along with Prisoner's Dilemma games.

Prisoner's Dilemma game is another milestone in ABMS field's development, and it represents a unique method of understanding human behaviour in critical situations of life. In this revolutionary social thought experiment two imaginary prisoners are offered two distinct options: to cooperate and help the other prisoner by remaining silent or to defect and blame the other one via confessing. The prisoners are unable to communicate in any way. The choices and the prisoners' penalties are as follows: if both of them stays silent, each serves one year. If only one of the prisoners betrays (defects), the other one is sentenced to three years. If both of them betray, each serve two years.

The original game's researcher John F. Nash predicted defection of both participants - a rational selfish agent's tactic. This argument seemed perfectly rational as well, since independent of other player's move, confessing is more beneficial than cooperation. However, associated further social studies discovered just the opposite, people assisted each other (Aumann, 1959). This result promised valuable insight into the evolution of cooperation among human societies and it was

Axelrod, who seized this opportunity with the field's newly developed tool, agent-based modelling. He introduced a tournament to enable various Prisoner's Dilemma strategies to compete with each other. The winning strategy was also the simplest with the shortest algorithm: beginning with cooperation, always repeat the opponent's last move, named *tit-for-tat*. This is the programmatic application of universally known strategy "An eye for an eye, a tooth for a tooth".

Tit-for-tat surpassed other strategies in many similar simulation tournaments, sometimes with some variations introduced by "genetic algorithms" (Miller, 1986). We, however, propose that this insightful game theory story still needs further study on, to get closer to the real human nature.

A one-to-one social relationship is exactly what The Prisoner's Dilemma Game models. People are randomly (later rule-based) assigned to each other in pairs of two. Our social daily life, however, mostly consists of interactions within a group. When we talk more than one person listens, when we assist the elderly many people appreciate this, and any aggression creates fear not only in the subject but also in the people nearby. This novel group gaming concept would introduce more than four outcomes: A prisoner may blame one, two or three... other prisoners in the group and their penalties will intensify. Anybody will develop a view on anybody else and these opinions will be affected by results in every round. This study might lead to determining the most rewarding strategy to prevail and lead, after joining a new group. We therefore suggest that further studies should concentrate on group gaming, rather than randomised two-opponent rounds in the iterated Prisoner's Dilemma Game.

ABMS field has contributed to, and benefited from computational social science development and the growth of modelling platforms. Numerous tools and programming language libraries have been developed, most of them employing object-oriented software paradigm. Resemblance between the computer science entity “object” and our modelling unit “agent” has facilitated this. Several of best-known platforms are:

Netlogo: An ABMS development environment constructed in the spirit of Logo programming language and involves the highest-level programming experience of all platforms (Wilensky, 1999). Despite its shallow learning curve, it has a unique programming language with unfamiliar syntax. Low performance in modelling, probably stemming from the interpreted nature of the language (Wilensky et al., 2006), and its restrictive nature are criticised by scientific user community (Abe, 2010). There are opposing studies as well, claiming efficient computing (Railsback & Lytinen, 2012),

RePast (The Recursive Porous Agent Simulation Toolkit): An agent-based modelling library for several programming languages, developed by Collier et al. (2003) at the University of Chicago. The original library was developed for Java but variations for Python, .NET and C++ coding environments are available. The library has been praised for high execution speed (Railsback et al., 2006) and comprehensive documentation (Getchell, 2008). It provides a middle ground between convenience and scientific functionality.

MASON (Multi-Agent Simulator Of Neighbourhoods... or Networks... or something...): This is an ABMS library for the Java programming language. Faster than RePast (1% to 35%) (Railsback et al., 2006), this library was designed with speed of execution in mind and it enables most efficient computing for models. However its lack of features and immaturity renders it unusable for most cases (Railsback et al., 2006).

SWARM: This library is Santa Fe Institute's product but it has not been maintained recently and seems to be outdated. (Latest stable release was in February 2005).

We have concluded that RePast excels at being the good compromise between functionality and simplicity and therefore RePast will be utilised for implementation of the model.

Wikipedia and Collaborative Content Creation

Wikipedia is the collaboratively created community-led free encyclopaedia that anyone can edit. It symbolises one of the most valuable outcomes of cooperative effort in human history (Yasseri, 2012). Its growth has been tremendous throughout this thesis' preparation period. In February 2013, it covered 24.6 million articles (Wikipedia:Statistics, 2013) and as of July 2014, that number raised to 32.7 million articles (Zachte, 2014) in 286 languages. This growth continues, with 14963 articles per day in July 2014 (Zachte, 2014). It is maintained by a productive community of 1889677 people (Zachte, 2014) including members who contributed more than one million edits (Titcomb, 2012). It surpassed any expert-driven encyclopaedia written to date (e.g. last printed Encyclopædia Britannica had 65000 articles (Flood, 2012). Accuracy concerns and academic unpopularity do not seem to have ground since for instance a 2005 Nature study found Wikipedia's accuracy rate approximately equal to that of Encyclopædia Britannica (Giles, 2005) and a recent academic study measured "better" article quality than Encyclopædia Britannica and a textbook, on mental health subjects (Reavley et al., 2012).

Wikipedia has become a paragon of collaborative production in modern era; it is researched in over 400000 of academic papers according to Google Scholar, and

125000 news stories mentioned about it in Google News. It shaped a new form of content creation that has been employed in diverse set of fields from organisational knowledge (i.e. knowledge management systems, the environments for experienced employees to cooperatively prepare training documents) to scholarly literature (e.g. Scholarpedia, Connexions). It creates a novel and unique reason to believe in the benevolent side of the human nature, as with no official compensation for contributors except for a few core maintainers, with no established administrative system, even with no requirement to disclose personal information; the system works. The system works for democratising information, approaching hot topics with no obvious bias in the long run, and low-cost access to whole body of information for everyone. Understanding this phenomenon will lead to discover other new areas of application for this altruistic collaboration (Kuznetsov, 2006), of supposedly democratic-anarchic nature (Muller-Seitz & Reger, 2010). From organisations training new employees to universities creating textbooks, to even states preparing a constitution people should find the ways of harnessing power of the crowd. This would result in a better organisation in Wikipedia itself – for instance we might point out which areas need editing or which projects lack adequate contributors. Although *wiki* (user-editable web pages) is a recent concept, with its first implementation *wikiwikiweb* going on-line in 1994 (Cunningham, 1995) and Wikipedia itself going operational in 2001, and it fuelled a rich literature ecosystem around the phenomenon.

In order to systematically examine and meaningfully explain the Wikipedia's working system Julien (2012) created a framework for collaborative content creation process. This study models online collaborative and productive effort as an independent system, consisting of inputs (both users/producers and environment), process (interaction and organisation) and the outcomes (product/information, structure, volume). Users' competency and motivation, as well as initial and ongoing investment make a starting point, whereas online activity, roles, governance provide the required work and process. Resulting product is the information itself, as well as its coverage and structure.

As an extension to this work, we propose that ongoing process (online activity/community) and collaboration are by-products as well, as we would not have online self-organised production model without them. Associated work of Julien (2012) is shown in Figure 3.

Wikipedia is proposed to resemble a living organism (Kamalabadi, 2006) which lived *infancy* in 2001 with explosion of unorganised superficial articles and stubs. In *childhood*, which covers 2002-2004, Wikipedia actively gathered any information independent of having ground or not. This period "is characterised by the organisation and polishing of articles. New articles still flourish, but they are of much higher quality". *Adolescence* was the period between 2005-2007 and mostly consisted of harsh debate and edit wars. This period witnessed a surge in general interest and wide adoption by the society and prevalent discussion assisted this penetration. Wikipedia then experienced a maturity age, *adulthood* from 2008 onwards. We suppose recent slowdown in article growth and edit count could be explained as the consequence of this. Not only did Wikipedia cover considerable amount of known science and encyclopaedic information but also its contributor community altered. This maturity might be thought of as an outcome of new users' perception. The latest Wikipedians are mostly observers since in their view, Wikipedia has always been there for them - it is taken as some form of an art product. In this sense, editing Wikipedia means like modifying the text of *Les*

Misérables. For Wikipedia to return to its previous days of high-performance growth, it needs the amateur spirit of the youth – both Wikipedians' and Wikipedia's youth.

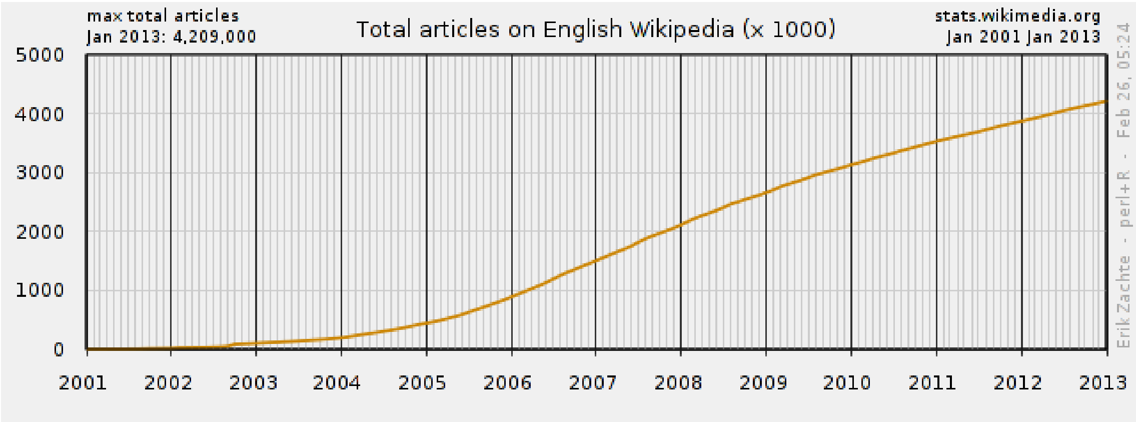


Fig. 2 Total article count in English Wikipedia.

A typical new language Wikipedia edition demonstrates a common growth pattern. Article count grows exponentially until maturity. Then a linear pattern is observed. The most established encyclopaedia edition - English Wikipedia's development timeline could be traceable in Figure 2 (figure obtained from http://stats.wikimedia.org).

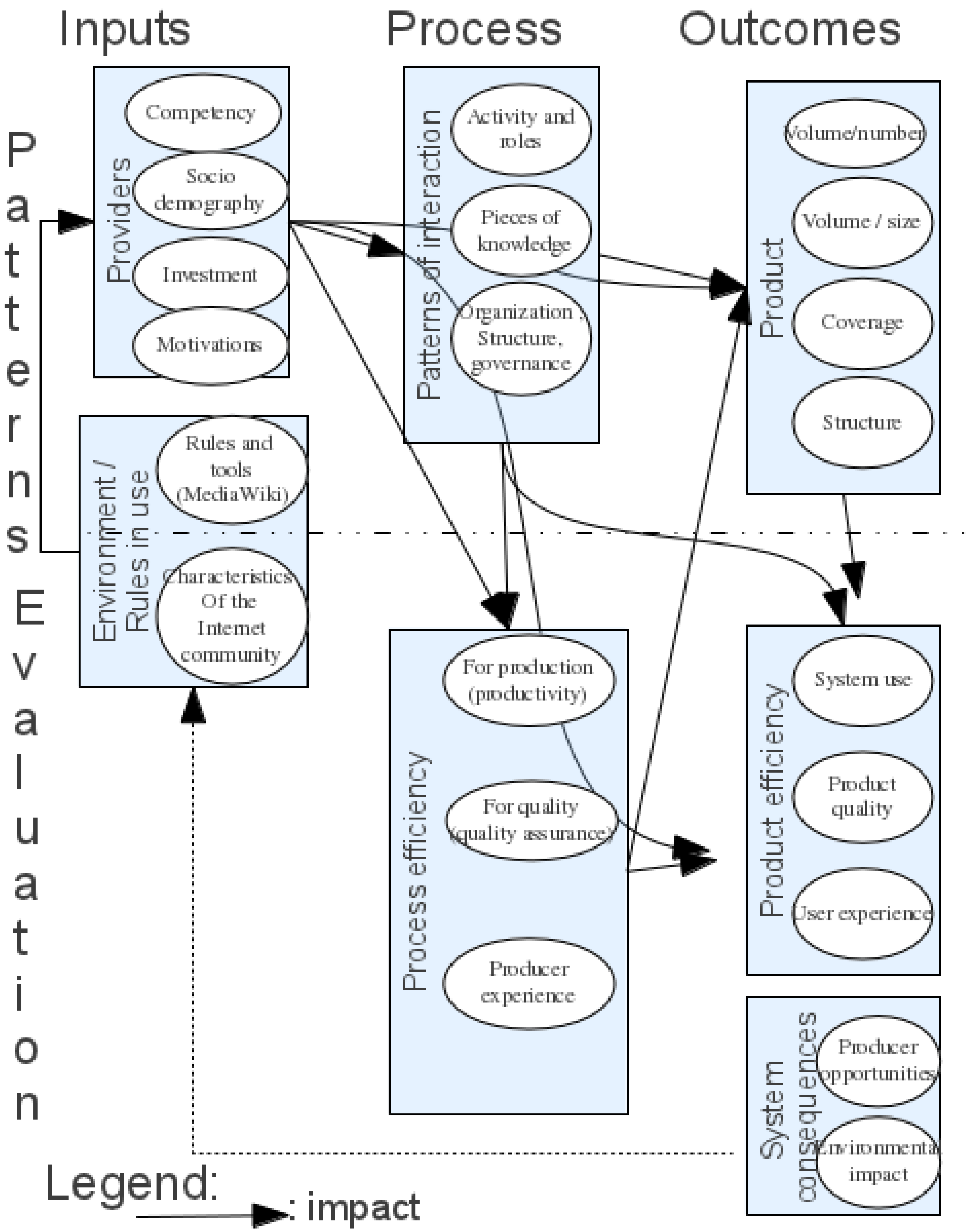


Fig. 3 Inputs, process and outcomes of online open projects (Julien, 2012).

This growth pattern has been modelled with *Gompertz function*, an empirical equation: (Wikipedia: Modelling Wikipedia's growth, 2010).

$$y(t) = ae^{be^{ct}}$$, with

a= 4378449 b= -15.42677 c= -0.384124

t is the time in years since 1/1/2000 (so 1/1/2010 is t=10.00).

However, this growth is the outcome of Wikipedians' work, and this study will model it accordingly, via agent-based modelling. Associated active editor population for English Wikipedia shown in Figure 4 points out a clear trend change in 2007.

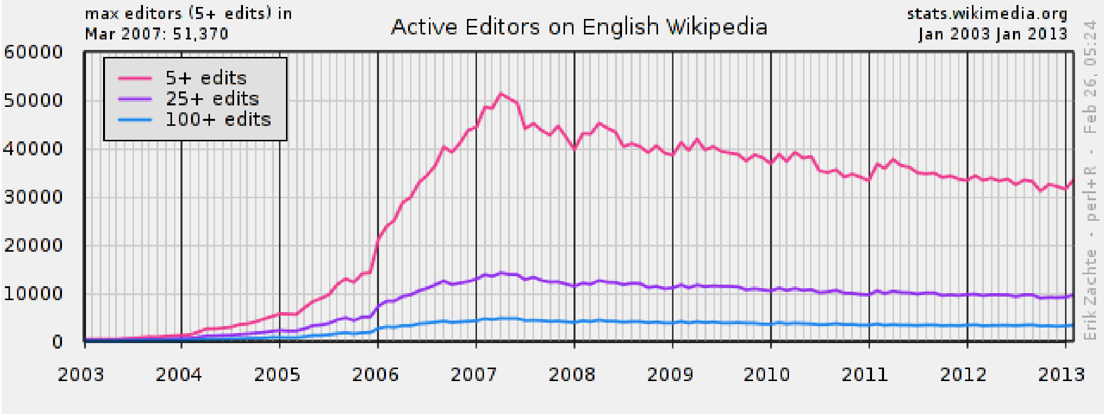


Fig. 4 Active editor count timeline in English Wikipedia.

An *active editor* is defined to be a registered (and signed in) user who made 5 or more edits in a month.

Pageview pattern in an established Wikipedia language edition (English Wikipedia) follows a slight linear growth, without an apparent slowdown (Figure 5).

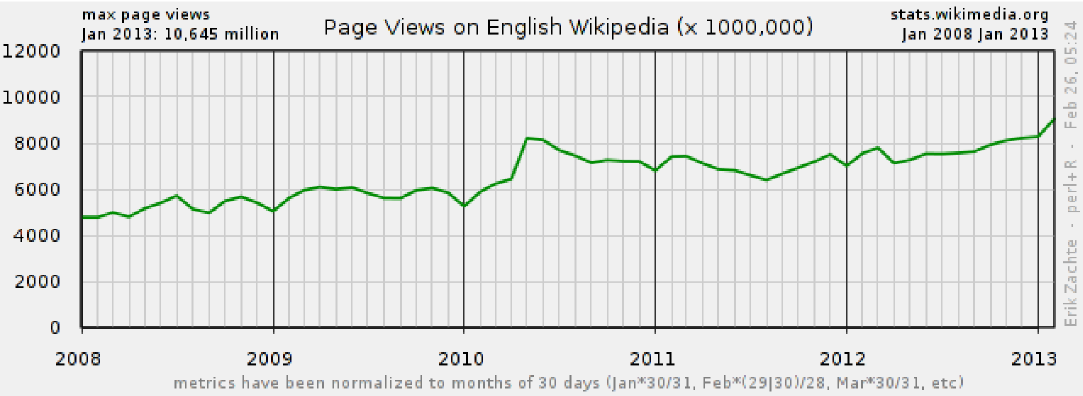


Fig. 5 Pageview count timeline in English Wikipedia

Modelling to Understand Wikipedia

Because Wikipedia is an output of autonomous and asynchronous-behaving agents (encyclopaedia editors and users) interacting with each other, it is a good example of emergence. It is difficult to predict that collaboration of non-professional internet-interested individuals might produce a massive encyclopaedia and a social phenomenon. With its community-centred self-organisation and self-healing capabilities, Wikipedia is a well organised representation of its users' and contributors' knowledge, thus, modelling it based on these users and contributions will approach its true nature better, bringing about more meaningful results (Ingawale, 2008).

Typical Wikipedians are our next focal point, as the realistic model of Wikipedia will only follow after a faithful model of its contributors. A typical Wikipedia.org visitor spends approximately five minutes per visit to the site and sixty seconds per pageview (Alexa: Wikipedia, 2013). Wikipedians (Wikipedia community members) are grouped into two basic classes by Anthony et al. (2005):

Good Samaritans and Zealots. Good Samaritans are passers-by in the online environment that make only one edit or rare contributions from time to time. They typically do not register and often do not disclose their identity. A surprising result is the highest quality contributions are supplied by this group: unique one-time edits of Good Samaritans are the most sophisticated and valuable ones. Zealots, on the other hand, dedicate themselves to the community often pursuing reputation and status. This group includes truly exceptional members, even an editor with one million edits. As the edit count/frequency increases Zealots' contribution quality increases, while Good Samaritans' decreases – this type of Good Samaritan is probably an offender vandalising the articles concealing their identity (Anthony et al., 2005)

Among academic efforts to examine social development process and community collaboration of Wikipedia, Ingawale et al. (2009) utilizes graph theory concepts and proposes "small worlds of Wikipedia", analysing the free encyclopaedia as a *small world* network. Graphs of this type include clusters (tightly connected group of nodes) while random edges between different clusters exist. Ingawale (2009) includes results of processing Cebuano language Wikipedia edit history, and interrelated users and articles clearly demonstrate a small world scheme. Obvious clusters begin to emerge as the encyclopaedia grows.

Viégas et al. (2004) adopt an article-specific approach and produce a tool to represent a Wikipedia page's activity, namely history flow visualisations. These visualisations point out several common patterns in an article's development process. Edit wars are apparent, while visualisations also stress powerful self-healing capability of Wikipedia.

Ingawale (2008) argues, and calls for exact subject matter our study in his 2008 conference paper, "Understanding the Wikipedia Phenomenon: A Case for

Agent Based Modeling". He states that this approach would help explain self-organisation in Wikipedia, while offering future opportunities to employ in knowledge management systems. Despite interest, such an implementation of an agent-based model for collaborative content creation or collective intelligence is difficult to find in literature. A recent master's thesis (Yıldırım, 2009) is one of few studies that aim to model wiki project on the behaviour of its users. This study claims to have found a relationship between the degree of centrality of a node (category in a wiki project) and the barrier of the category (defined as approval and rejection count of the category).

CHAPTER 3

THE AGENT-BASED MODEL OF WIKIPEDIA

Why Should We Model Wikipedia?

Before constructing a meaningful and faithful model of Wikipedia, motives and potential benefit of such a model needs to be questioned. Why should Wikipedia be modelled? Beyond being an online encyclopaedia, Wikipedia is a prototype for the success of decentralised collaboration. It is a new form of online interaction, a new form of communication, a new form of content creation, among others. It is particularly valuable for its outcome as well – a massive encyclopaedia, images, videos, news, database constituting an open body of knowledge on an unprecedented scale. Additionally, content creation is required in other areas, which would be apparent application area candidates of this remarkable system. From academic papers to internal training documents for companies, this model of interaction might be utilised for a more efficient, effective and economical production. Wikipedia is additionally a chain of trust – many strangers work together to build it. This communication-less cooperation is of particular interest as well, which may find itself many other fields of practice.

Wikipedia's success factors might be multi-dimensional. Previous efforts to cover all human knowledge can be traced back to the onset of French revolution, *l'Encyclopédie* (1751), however collaborative effort of French revolutionaries and later efforts lacked Wikipedia's main instrument: the Internet. This is the timing factor. Additionally, Wikipedia is an organised body of meaningful *information*. Information or one's knowledge can not be removed deliberately, it increases when shared, and more importantly, everybody has some degree of knowledge. This raises interest and removes barriers for joining the circle. Information has a very low transaction cost to edit as well. If it were a physical equipment to be co-built, every contributor would need to send the last version to the next – a costly shipping. Information, on the other hand, is relatively easy to store and takes milliseconds to send.

Wikipedia's model of collaboration has the democratic approach in itself, raising its importance. Every individual has the right to speech and contribution, but eventually community determines which ideas will survive: the democracy. Therefore democratic societies might benefit from Wikipedia's experience as well, by opening any laws or regulations to active collaboration for editing after the initial work. This would enable ordinary citizens to participate in the legislation – the ultimate goal of any democratic regime.

Despite its exceptional mechanism, and structure worthy of analysis, Wikipedia is actually difficult to experiment with. Not only are many users anonymous, scattered around the world and hard to reach, but also Wikipedia's very nature – decentralised and chaotic collaboration – prevents top down approaches for analysis of the working system. As there are no executives or no main owner, the only opportunity is analysing Wikipedia in a bottom-up fashion. The focus needs to be on the ground, typical contributors and workers of the free encyclopaedia – its users. This is in fact the most basic rationale behind agent-based modelling: When the system is too complex to understand, concentrate on the meaningful basic units – agents.

In addition, such a model would contribute to better organisation in Wikipedia itself. Is it beneficial to encourage users to reveal their identity; should editors focus on their expertise or contribute in diverse areas as well? If answered, these types of questions would greatly assist in the foundation of a new language edition of Wikipedia. Its *administrators*, *bureaucrats* would act and maintain according to lessons learned in such an experience.

As the motives and instrumentality are clear, aspects of an intuitive and productive model should be determined, beginning with an agent candidate. In Wikipedia ecosystem there are two main objects to consider as the agents: articles (including edit history pages, project management pages, etc.) and users (readers, writers, editors, administrators and all sorts of Wikipedians). Both of these agents are autonomous and interlinked in their environment. They are “autonomous” in the sense that neither users nor articles are triggered or removed by actions of other agents. They are interlinked, as articles link to each other and the users are co-workers of this system. Although articles are not intuitive to be treated as the model’s active agents, as articles cannot act or develop on their own, and that is the users who are independent decision-makers behind the development of such a system; our model treats articles as agents as well. They are passive agents which users edit, therefore connect to. Additionally, we take relation between articles into account in this editing activity. We take same user's editing activities in various articles as interlinking of those article agents as well. Therefore a user's contribution results in three major types of connections in three separate networks:

1. The user agent connects to the article agent in the general *Affiliation Network.*
2. The articles, which the user previously edited, connect to the new article in the *Article Network*.
3. The user connects to previous authors of the article in the *User Network.*

These networks' structure and the model's mechanism will be explained thoroughly in the following chapters.

In order to examine the interactions among users (main agents), it is required to define the meaning of links between them, and how they are building connections. Because Wikipedia, and particularly its articles, are the main point of interest, we should depict the interpersonal bonds accordingly. If a user creates a *stub* (the Wikipedia term for undetailed, immature articles) and another user edits it and adds better information about the subject, these two users, independent of whether they know each other or not, are linked. They “cooperate” to write an article. This way, every contributor of a particular article (over one thousand for many articles) belongs to a group, or network, of interlinked individuals.

Our basic agent, a Wikipedia user, has several behavioural templates that they may follow when they contribute to Wikipedia. A user can only do four things with an article: They may create a new article, read an existing one, edit it, or delete it. Reading is a passive act and does not lead to a connection to other users, in our model. The other three, on the other hand, makes the user more tightly connected to the online community.

A Wikipedia Graph?

Graph theory is an abstraction of real life, which includes entities (nodes or vertices) connected to each other with lines (edges or arcs). It has been firstly developed and proposed by Euler (1741) and generalised by Cauchy (1813). Graph theory finds numerous application areas from social networks to geographical positioning services.

Wikipedia is neither a map nor a social network so it is not intuitive to represent it as a graph. It is just a well-organised set of articles. This is, however, just tip of the iceberg. Only roughly 14 per cent of the pages in Wikipedia are read and edited – articles (Wikipedia: Statistics, 2013). Bulk of information lies within edit discussion pages, project organisation pages, user profiles and so on. Wikipedia, although not a hierarchical environment, is an outcome of a tight connected network of capable individuals. Wikipedia community members specialise in their area of contribution (locating in different points of infinite knowledge plane). Users with similar interests communicate and cooperate for content creation. Because network structure is not apparent, we need a method to generate a graph consisting of vertices and edges (or a network). In our model, if a user edits a page he connects to all other previous editors of the page, whereas later edits of a user on other pages connect first edited page to all others.

Hypertext Transfer Protocol (http) that connects information as well as people on the Internet offers a rather obvious means to be represented by a graph. Web pages point each other with one-way directed arcs: hyperlinks. Wikipedia is a prototype of the Internet cloud, and includes excessive use of hyperlinks for both development (red internal links of Wikipedia for non-existent articles or *'citation needed'* tag) and navigation (blue internal links of Wikipedia or links for external web pages).

When we represent web pages as the nodes and hyperlinks as the directed edges, we obtain a highly asymmetrical graph. Some pages belong to the privileged set, accepting majority of the links, whereas rest of the cloud consists of web pages with only several links. If we draw a web page count - link count graph, we observe an inverse proportion pattern, with a *'long-tail'* of vast majority of insignificant web pages. This pattern is observable in Figure 6 (Adamic & Huberman, 2002).

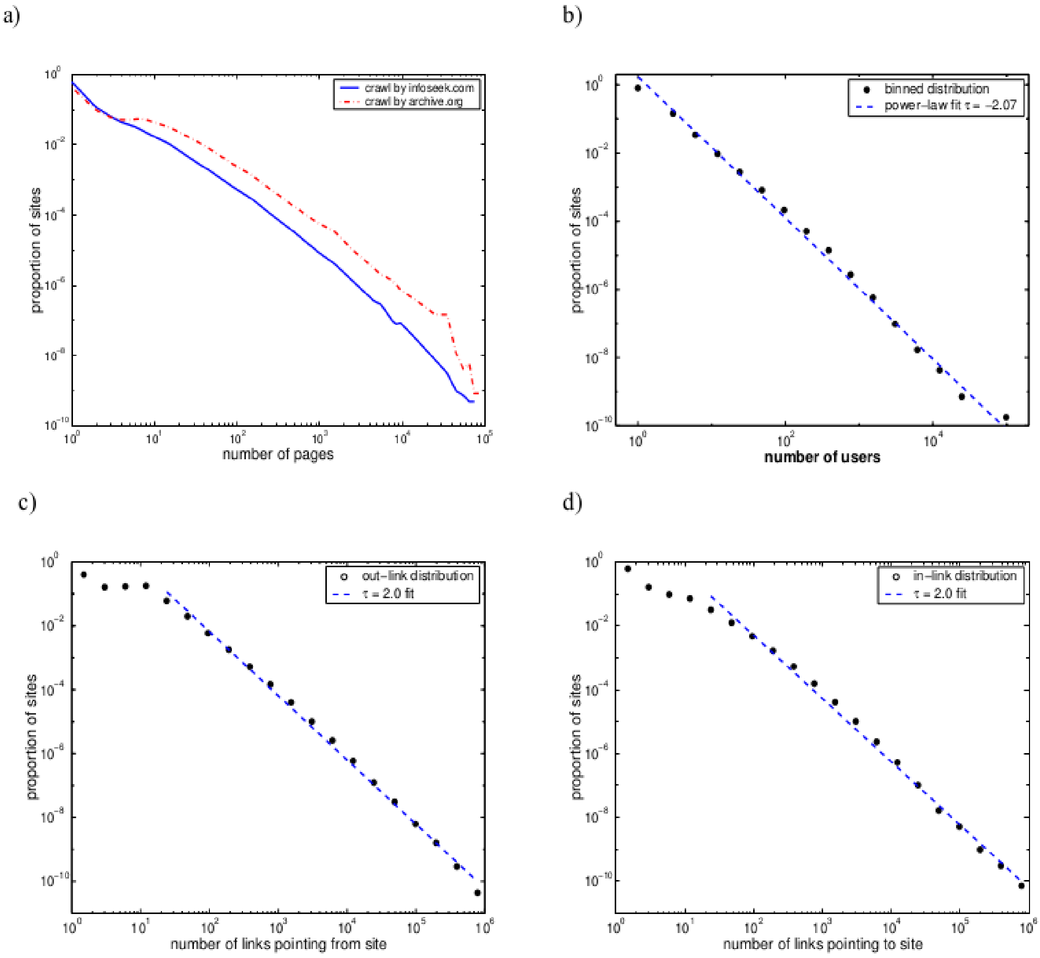


Fig. 6 Fitted power law distributions of the number of site a) pages, b) visitors, c) out links, and d) in links, measured in 1997.

One particular class of graphs especially useful for internet-like networks and this class is named as *small world graphs* (Watts & Strogatz, 1998). We follow a certain procedure to reach a small world graph: we begin with a neighbour-connected graph and with probability p we break a certain nodes neighbour connection and connect it to a random other node in the network. This model of a graph-based network captures clusters in online environment but requires refinement for better modelling, and an extension of it will be utilised in our model. This extension is a special type of *preferential attachment* (Barabasi & Albert, 1999) in which current condition of vertices affect probability of future connections. As the current degree increases for a node, probability to have more new edges also increases, creating a few *hubs* in the graph with substantially more connections than other nodes.

We propose a more central approach for small worlds. In online environment small worlds (clusters) exist, in our view, but they are not randomly scattered around the network. Tightly connected nodes (ones with more edges connected to them) are more likely to belong to a cluster. This is because clusters (small worlds) emerge around them. If a highly edited article on Wikipedia is taken, like "astronomy"; it is more likely to belong to (actually create) a cluster than the article about a newly discovered asteroid. Therefore we slightly alter small world graph procedure: after break the link to a neighbour with probability p, rather than randomly selecting a node for a new connection, we limit the pool of candidates for edge destination, to nodes with large number of edges (e.g. those nodes which have more edges than two times the average degree of nodes in the network).

Because we postulate functional resemblance between Wikipedia and the Internet as a whole, we expect to see similar pattern of long-tail in Wikipedia articles as well. Wikipedia *main page* is the concentration point of internal and external links whereas newly created stubs would not connect with that many links, and stub count far exceeds quality article count. Two thirds of the articles were stubs in Wikipedia in 2010 (Gray,2010).

We should detect this type of long-tail pattern in user contributions as well, since more contribution means better quality article (Wöhner & Peters, 2009). Users should gather around some central subjects, although outliers exist as well. This leads us to the notion of general contributor traits among Wikipedians. Some prefer one-shot contribution to create a new article and forget about Wikipedia (Good Samaritans) whereas others try to gain reputation and status in online community via constant contribution (Zealots). Furthermore, these natives of Wikipedia, zealots, differ among themselves in their frequency to edit/contribute, and their career choices (editing many diverse articles or focusing on their point of interest). The described connection building method, and basic character traits of Wikipedians constitute the fundamentals of our model and graph, for the online encyclopaedia community.

The Model: Agent-Based Model of Wikipedia on a Graph

Wikipedia covers a massive amount of information categorised in most of scientific/social/academic fields known and aims to span the entirety of human knowledge. In this sense its enlargement area potential is unlimited. Additionally, it is not a static project, and is developing constantly. This is a striking fact, since while a measurement process for a characteristic of Wikipedia continues; its scope extends to cover new information, invalidating the result.

Taken all explained aspects into account, Wikipedia's body of information might be best represented as a *cartesian plane*, constructed in a nominal sense. This would be an unbounded plane of information organised in such a way that there are regions for every subject of interest of editors. For instance, the “Mathematics” region covers a large area and “Complex Numbers” is a subdivision of it. Articles are represented by “dots” in this cartesian plane of information, as well as contributors of Wikipedia, only with a different shape and colour. These dots locate in the plane in such a way that more relevant articles, and users with similar interests, are closer to each other. This is intuitive, as two mathematicians, for example, would probably contribute more about mathematics, and therefore would locate in the mathematics region, making them closer than a third physicist.

Higher or lower coordinate values do not have a comparative meaning in this plane; these are merely unique positional values. This type of abstract expression of Wikipedia is observable in Figure 7.

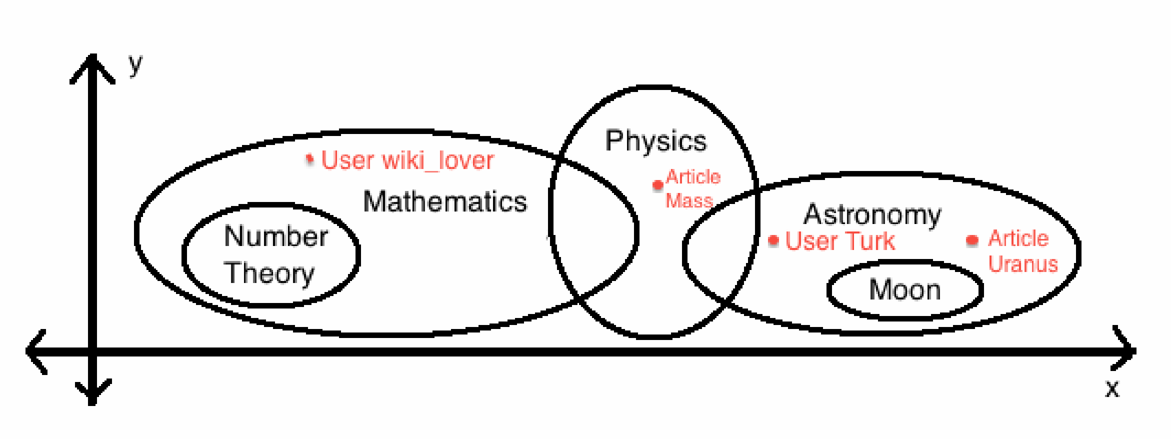


Fig. 7 Cartesian plane of information of Wikipedia, modelled in nominal sense.

Our model implements the described cartesian plane. Our plane is bounded and the number of agents (users and articles) is less than that of Wikipedia for the sake of feasibility and parsimony. Additionally, using a floating point for exact positions of agents would add far too much complexity to calculations. The solution we come up with, is to overlap this cartesian plane with a lattice of the same dimensions. This enables usage of integer number coordinates for agents (in addition to floating point exact position coordinates) and enables better measurement of neighbourhood dimensions (neighbourhood concept will be described in detail).

This cartesian-plane-with-lattice based ABMS case basically includes repeated re-processing of all agents (either a user or an article) in a meaningful time unit or iteration called a *tick*. Initially, users and articles are unconnected vertices on the cartesian plane. In every tick, user nodes build edges towards article nodes in the affiliation network, with every connection meaning an edit.

In an N * N 2D square lattice/grid spread onto same dimensional cartesian plane, every cell is initially empty - no one is using Wikipedia at first. According to a user defined parameter (i.e. user count and article count) we generate agents and locate them in the plane randomly (randomizing operations are based on uniform distribution). The environment on which agents are created is a double-layered two-dimensional plane. The first layer is a continuous plane where agents are produced with random floating-point coordinates. The second layer is a grid (lattice) consisting of one unit square cells and this layer is used to calculate neighbourhood dimensions. Generation of a user node means a new user joins the circle, whereas an article node represents a subject proposed to add information. But agents have not contributed yet – only reading contents or developing interest for Wikipedia. In addition, for now, articles are only titles. At the end of this initialisation phase we have userCount (default value 1000) interested individuals ready to contribute about articleCount (default value 1000) topics scattered randomly around the lattice.

Table 2 Main User Types in the Model

User Type	Career Choice	Characteristics and Contribution to Wikipedia
Good Samaritan	-	Contributes in the first tick only, by adding information. Might be anonymous.
Zealot	Project Leader	Contributes in every tick, by adding or editing information in their area of expertise (i.e. a neighbourhood in the plane), for reputation.
Zealot	Administrator	Contributes in every tick, by editing *Good Articles* for grammatical errors, or according to guidelines, for reputation.

The process of connecting the user-article graph differentiates according to characteristics of the agents. Rather than a monotonous single type of agent, our model points out and simulates three distinct Wikipedia user styles. Users may prefer to contribute only once, concealing their identity. These are named *Good Samaritans* in Anthony et al.'s (2005) study. They are passersby in the online community, adding a major body of information in a few edits (exactly one for our model for the sake of simplicity), often not bothering to create an account or to pursue an image or status. The next choice is turning into a *Zealot*. Zealots are the natives of the online environment. They are after having a name and status. They have personal accounts, user pages, and they join in decision-making process for the free encyclopaedia.

In this step, when determining the roles of users, we first check simulation-operator-defined parameter values about contributor types. The *Active user percentage* parameter determines how many of the users will be created as zealots to make systematic contributions and how many of them will be created as good samaritans who stop after first edit. This characteristic of agents is all-inclusive and mutually exclusive, i.e. every agent is either zealot or good samaritan.

The third step focuses on career preferences of zealots. Starting from cell (1,1) if a cell contains one or more zealots, we determine a career path for them. There are mainly two career opportunities. According to an operator-defined parameter, *General Interest Percentage,* they may prefer to edit the diverse set of *Good Articles* (e.g. they may correct grammatical errors in articles with a substantial number of edits). This choice ultimately takes them to the *Administrator* position. Otherwise, they may concentrate on their area of interest and this path takes them to the *Project Leader* position eventually, as shown in Table 4.

The fourth step is the beginning of collaboration, and the next steps repeat the same algorithm (with some derivations) over and over. The fourth step handles good samaritans' one and only contribution and utilizes the *Neighbourhood Connection* algorithm (which will be explained in detail). For every cell that includes a good samaritan, we use a neighbourhood connection algorithm to create an edge with another agent in the vicinity of the cell. Because in the continuous plane a region/neighbourhood represents an area of interest in a nominal sense, connecting to a nearby article node means adding information to, or editing the article in question, cooperating with agents having similar interest (users having an edge to the same article node). This represents a user (good samaritan) writing or editing an article in his area of interest. This iteration generates the first layer of connections and good samaritans will not get connections anymore. They will simply be passed unchanged for later iterations as they are assumed to contribute only once.

Beginning with fourth step, in every agent connection operation, a *Good Articles* array will be updated. “Good article” label is a quality indicator for articles by which Wikipedia adjusts its approach towards them. They are grouped into a specific section and offered as the best of Wikipedia, which lead to a peak in community interest. In our model, good articles are the ones that are regulated by the edits of “administrators”. Although the real procedure for selection requires candidate status and community consensus for promotion (“Wikipedia: Featured article candidates”, n.d.) we utilise a highly-edited-means-high-quality approach as there are findings on this (Ingawale, 2009) and in this thesis we argue for the adoption of an easier and more automated selection process. The good article array keeps a window of article nodes with most connections (the ones whose edge count is greater than a parameterised multiple of average edge count of articles in the network). Every new edge generation may modify included articles, increasing the connection count for an article by one. When a new article is added to the list, the first member is excluded from the array, preventing excessive processing of articles added earlier.

The fifth and later steps process grid cells containing zealots. For every zealot, if they are a project leader candidate, we apply a neighbourhood connection algorithm to create an edge with a nearby article. Otherwise, they are on the administrator career track and we utilize the *active agent connection* algorithm (will be explained in detail) to connect him with a member of good articles array. This metaphorically represents a general edit (e.g. correcting the date of Turkish Independence War) in a real life online community. Users who prefer to edit articles on various subjects tend to apply Wikipedia guidelines for these articles, eventually raising the user to the administrator position in the community, while also helping the article for receiving "featured" status.

Beginning with the sixth step, the model basically connects active agents (zealots) to related article nodes, according to their interest (location on the plane) or motivation (The active agent connection algorithm raises good articles' edges more, as a side effect. This leads to gap between elite and normal articles of the encyclopaedia, in a "rich gets richer" manner. This trend effectively captures the real life phenomenon of long tail in scale-free graphs like the Internet).

We repeat the operations of the fifth step until a pre-determined tick/turn/step count is reached, or until the system stabilises. We expect the resulting graph to include both small-worlds/clusters (through the neighbourhood connection algorithm, distance is inversely proportional to edge count) and the asymmetrical nature of the world wide web, the long-tail phenomenon (through good article connection algorithm, top article nodes build the majority of the connections, and the rest constitute the tail).

Neighbourhood Connection Algorithm: According to *neighbourhood* *dimension (ND)* parameter specified by the operator of simulation, we take the Moore neighbourhood of diameter ND around the agent in question, we scan for finding an article to edit, in randomised order. Then we repeat the process for two outer rings (the second circle with twelve cells, the third one with sixteen cells). This first step is *the neighbourhood search algorithm*. If an article node is found in the neighbourhood, we create an edge between it and the aforementioned agent, and we stop the process. If no article agent is found, we stop the search and jump to the next cell for associated agent operations.

Active Agent Connection Algorithm: For the agent in question, we look up the good articles array, shuffle it, randomly choose one article node, and create the edge between the result and our agent.

In first draft of the model, we designed the graph consisting of only users, representing only the collaboration between the Wikipedians. As the next version, the model has been generalised to include articles on the graph. This addition results in a heterogeneous graph (including both agents and articles) and increases the complexity as well as dynamism and closeness to reality. The algorithms have been modified accordingly. In this version, edges connect two different entities, agents and articles, making them arcs (one-to-one directed relationship). This type of edge better simulates observable phenomena, describing a direct relationship between nodes (e.g. a user *edits* an article). Throughout the process we develop a multimodal collaboration network graph, which consists of articles and users. This type of heterogeneous graph structure is referred to as an *affiliation network* in the literature (Lattanzi & Sivakumar, 2009).

Because fundamental graph theory metrics such as average clustering coefficient and path length of a network is defined for unimodal homogeneous graphs, the affiliation network needs to be projected distinct user and article networks. The method, introduced in Ingawale et al. (2009) and employed in this study, is executed as follows:

1. If user agent A currently edits an article agent B, connect user A to users who previously worked on that particular article in the *user network*. Similarly, next users editing the article will connect to agent A as well as previous editors.

2. If agent-user A currently edits an article B, connect article B to previous works of agent A, i.e. articles edited by Agent A. These connections are built in the *article network.*

Consequently, we have an affiliation network graph consisting of both articles and users, and two unimodal graphs, which consist of projections of the general network into article-article and user-user graphs.

Our social simulation experiment's ultimate target and outcome is to determine the relationship between our model's designed parameters and the resulting graph's principal characteristics. This is illustrated in Figure 8.

Model Inputs	Mediating Parameters	Outcome Characteristics
Cartesian plane of information	Active user percentage	Total edge count
Grid of neighbourhood	General interest (Admin) percentage	Path length
		Clustering coefficient
Article nodes	Neighbourhood dimension	Small world characteristics
User nodes	Good article multiplier	Degree distribution
	Good article connection count threshold	Degree correlation

Fig. 8 Basic input-output variables and mediating parameters of the model.

Wikipedia, in our model, is the outcome of dynamic interactions among its users. This massive network of collaboration needs an initial environment and inputs; a process to utilise those inputs; and the resulting product, the graph (the network) and the encyclopaedia. The basis of all the inputs is the universal human knowledge, which is represented as the cartesian plane of information. Other requirements are the interested individuals to add this body of information to the encyclopaedia (user nodes), and the structured subjects to cover (article nodes). These phenomena constitute the model inputs.

Mediating Parameters and Output Variables of the Model

In the process of collaboration, several parameters are thought of as the factors mediating the activity. These parameters are described below:

Article/User Count: These variables are set by the operator prior to the beginning of the simulation. In this way article count represents possible subject areas, or stubs; whereas user count determines how many individuals are interested in a newly founded edition of Wikipedia. As new connections i.e. new edges pointing to the articles, are created by the users/editors, stubs evolve into established, featured articles.

Dimensions of the Environment: As one of the most fundamental properties of the graph, this parameter mediates the density of the network. Larger environments, i.e. sparser networks, reduce the chance to find a peer agent in the neighbourhood, which may result in unconnected sub-networks. In our model width and height of the environment can be set unequal to build any kind of rectangular shape, but still, this parameter is a limitation of the simulation. This is because the total knowledge of humankind is unbounded, and Wikipedia grows to cover a larger portion of it every second. Future studies might expand on this to include an enlarging environment.

Active User Percentage: This parameter determines the “zealots (active users) / total users” ratio. This means this active group contributes every tick (turn) whereas the rest are good samaritans adding new content only once, in the beginning of the simulation. This value is adjusted so that active users are the minority in our model, in parallel with the real situation in the free encyclopaedia.

General Interest (Admin) Percentage: This represents the proportion of zealot users who are on the administrator career path, which means they are able to connect to “good articles” (articles with considerably high edit count, previously described) with every tick, independent of distance in the cartesian plane. These users make general edits to established articles

Good Article Connection Count: This parameter sets the minimum number of connections (edges) a particular article needs to have in order to be checked for good article status. This prevents good article selection in the early stages of the encyclopaedia.

Good Article Multiplier: This value directly affects the good article status of any given article. In our model, a good article is defined as a member of the highly-edited, most contributed group of articles whose edit count is greater than a minimum value (*good article connection count threshold*) and it is greater than Good Article Multiplier *(times) Average Connection Count of Articles in the network. More edits means more work to create a specific article, which is thought to be of better quality compared to others in the system.

Neighbourhood Dimensions: This parameter metaphorically represents the area of Wikipedia users' topic of interest, e.g. a biology expert may restrict their edits/contributions to only biology topics, or they may prefer to add content about medicine or genetics as well. This value determines the diameter of the Moore neighbourhood around the agent of interest, so if value is set to 2, the user agent may connect to a node in a 5x5 square area around themselves.

These mediating parameters are thought to actively affect the quality, organisation and characteristics of the resulting network. This productive community network is represented as a graph in our model, and these properties of the network will be examined in measurable characteristics of the graph, as follows:

Total Edge Count: An edge means a unit of collaboration between two entities, in our model. Total edge count is therefore the most basic measurement for collaborative effort. Additionally, in the Affiliation Network, total edge count means total edit count in the encyclopaedia, since an edit is represented as an edge between user and article in this network.

Small-world characteristic Q: Small-world graphs have relatively recently gained interest in academic circles and are a middle form of random and fully connected graphs. This characteristic is a measure of how cluster (tightly knitted group which is weakly connected to outside nodes) oriented a graph is. The procedure for creating a small world graph is as follows: a chain of nodes in which every (k)th node is connected to (k-1) and (k-2)th nodes, and last node is connected to the first node, which is probabilistically 'randomised', i.e. for probability p an edge is removed and directed to a random other node. For p=1 a random graph is built and for p=0 a chain of nodes is kept. This characteristic measures how clustered a network is.

Path length: For a certain node n1, the path length is the distance of a random node n2 to it. The path length of a graph is the average distance of random nodes to every node in the network.

Degree Distribution: This characteristic of a graph identifies equality (or lack thereof) of edge count among nodes. If influential or active members exist in the graph, their degree average increases and the distribution leans towards them. The demonstration of the degree distribution is a column chart with the edge count on the (x) axis and associated vertex count on the (y) axis.

Degree Correlation: This parameter is a representation of the preferential attachment of high degree nodes to other high degree nodes. This is measured as proportionality of degree of a node and its neighbours for every node in the graph.

Average Edit Count For Articles: Edit Count (represented as degree of an article node) measures contribution efforts of an article and is the basic quality indicator of it in our model. The average edit count for articles, therefore, is our main tool for overall quality of the encyclopaedia simulated. We use this indicator for assessment for success.

The explained five mediating parameters are examined for their effects on the resulting six graph characteristics. A simulation replication is run first for the default values for the mediating parameters, and then every parameter is given a higher and a lower value. This way, 2x5=10 replications are run for the simulation experiment. Along with the base run and eventual need for one more experiment for the good article multiplier parameter, twelve experiments have been done. Succeeding sections will elaborate on this simulation process.

RePast library functions enable visualisation of graph appearance and agent activity.

The Explanation of the Algorithm Used in the Experiment

The simulation experiments have been executed via an algorithm in the Java programming language and the RePast simulation library for agent-based models, in the pure object-oriented fashion. Presented below is the pseudocode transformed into procedural format for readability. Basically, a typical run of the experiment consists of two initial phases, a main loop iterated until the end of the simulation, and calculation procedures for model data output. These are as follows:

1. The initial environment is constructed (same dimensional cartesian plane and lattice surface are overlapped; three separate networks, the main affiliation network for editing activity, the user network for user-user interactions, and article network for article-article interactions, are created; and initial values for input parameters are set).
2. User and article vertices are created and randomly located on the cartesian plane according to a specified amount, and user vertices are given roles as good samaritan, project leader or administrator according to specified ratios.
3. Good samaritan users make their one and only contributions, i.e. their vertices are connected to a random article vertex in their neighbourhood (an edge is created between and a closely located article, in their area of interest).
4. In the main loop, in every iteration (tick), every active (zealot) user makes an edit (connects to a article vertex) according to their role. Administrator users connect to good articles, whereas project leader users connect to articles about their expertise (nearby article vertices,

in their neighbourhood).

5. After a specified number of ticks, graph characteristics are measured: average degrees of articles and the small world characteristics as a assessment of quality, degree distribution and degree correlation as the analysis of the dynamics of collaboration, and the path length and size of the giant component as to understand how closely knitted the resulting community is.

The associated procedural pseudocode and pre-post conditions of the program's components, executed for simulation are shown in Figure 9, 10, 11, 12.

pre: empty cartesian plane of A x A dimensions exists - double array plane[A][A] = 0
AUP(Active User Percentage) is set by the operator
GIP(General Interest Percentage) is set by the operator
post: affiliation network before ticks

```
void initialisation(){

    FOR i = 0 to userCount
        SET j to random integer value between 0 and A
        SET k to random integer value between 0 and A
        Add user US to plane[j][k]
    END FOR

    SET p to random decimal value between 0 and 1
    IF p less than AUP
        SET US as project leader
        IF p less than GIP
            SET US as administrator
        END IF
        ELSE
            SET US as good samaritan
    END IF

    FOR i = 0 to articleCount
        SET j to random integer value between 0 and A
        SET k to random integer value between 0 and A
        Add article AR to plane[j][k]
    END FOR

    FOR X = 0 to A - 1
        FOR Y = 0 to A - 1
            IF plane[X][Y] includes a good samaritan
                CALL neighbourhoodConnection(X, Y)
            END IF
        END FOR
    END FOR
}
```

Fig. 9 Pseudocode of initialisation - deploys articles, users, connects good samaritans
pre: double array plane[A][A] exists with good samaritans connected

GACCT(Good Article Connection Count Threshold) is set by the operator
GAM (Good Article Multiplier) is set by the operator
post: affiliation network with all connections done

```
void main(){

    FOR i = 0 to tickCount

        FOR X = 0 to A - 1
            FOR Y = 0 to A - 1
                IF plane[X][Y] includes a project leader PL THEN
                    CALL neighbourhoodConnection(X, Y, PL)
                ENDIF

                IF plane[X][Y] includes an administrator AD THEN
                    CALL activeAgentConnection(AD)
                ENDIF

                IF plane[X][Y] includes an article AR THEN
                    IF edge count of AR is greater than GACCT
                            and GAM * average edge count of articles THEN
                        Add AR to GoodArticles[] array
                    ENDIF
                ENDIF

            ENDFOR
        ENDFOR

    ENDFOR
}
```

Fig. 10 Pseudocode of the main program - connects zealots to articles until the end of the simulation.

pre: ND (Neighbourhood Dimension) is set by the operator
post: User is connected to the first article found in the neighbourhood

```
void neighbourhoodConnection(double apsis, double ordinate, user US){

    FOR X = apsis - ND to apsis + ND
        FOR Y = ordinate - ND to ordinate + ND
            IF plane[X][Y] includes an article AR THEN

                Create an edge between AR and US

                FOR every previous editor ED of AR
                    Create an edge between ED and US in User Network
                ENDFOR

                FOR every previously edited article PRE of US
                    Create an edge between PRE and AR in Article Network
                ENDFOR

            ENDIF
        ENDFOR
    ENDFOR
}
```

Fig. 11 Pseudocode of neighbourhood connection algorithm- connects users to articles in their neighbourhood.

pre: GoodArticles[] array is not empty
post: Administrator user is connected to a random good article

```
void activeAgentConnection(user US){

    Shuffle GoodArticles[] array
    Take last member LS of the array

    Create an edge between LS and US

    FOR every previous editor LS of US
        Create an edge between LS and US in User Network
    ENDFOR

    FOR every previously edited article PRE of US
        Create an edge between PRE and AR in Article Network
    ENDFOR

    Remove LS from the array

}
```

Fig.12 Pseudocode of active agent connection algorithm - connects administrators to good articles

Note: Pseudocode notation standard used above may be accessed in http://users.csc.calpoly.edu/~jdalbey/SWE/pdl_std.html

Instrumentation and Parameter Variations

Before starting the simulation experiments, it did not take a long time before it became apparent that a desktop computer (2.0 GHz dual core processor with 2 GB RAM) would not supply adequate computational power, even only for the base run. After one and a half hours of iterating the simulation, the computer was just in the 60th step of the first replication (base run includes 20 replications).

Therefore, a division of the computer lab (The Master's Lab of Boğaziçi University MIS Department) was arranged for the experiment. The division consists of 11 desktop computers of identical hardware specifications, as follows:

3.2 GHz processor with 4 cores

4 GB random access memory

500 GB hard disk drive

Windows 7 operating system

These eleven computers executed simulation experiment's code for a week (7 days) in a continuous fashion. This seemingly long period of time might be confusing, but the complexity of the model rises with higher parameters, and general-purpose desktop PCs were utilised. The process of agent-based modelling and simulation is comparable to other complex studies such as sequencıng in bioinformatics, in which performance-intensive processing units are used (e.g. Taşan (Lohmann et al., 2012) utilised an eight-core unit with 16-GB memory). The main computer executed *base run* whereas others processed deviations in the parameters (5 parameters with a higher and a lower value; 5x2=10). The computers ran 4 separate processes of 5-replication experiments in order to harness the full power of quad-core processors.

Our simulation begins with the "base run" experiment, which would be used for future reference to measure the effects of modifications in the parameters. This first experiment consists of n=20 replications in which all the parameters were set to a constant value. The logic behind these assigned values is described below:

Grid Height and Width: A 100x100 unit cartesian plane has been employed, which both enables perception of distinct 2000 nodes when visualised, and facilitates efficient processing being not too large.

Article and User Count: 1000 Users collaborating on and editing 1000 Articles is sensible for a recently founded edition of Wikipedia, and Ingawale's paper (2008) examines Cebuano language edition, and both article and user count is below 2000. This values are suitable for a Knowledge Management System of a middle-sized company as well, 1000 employees might be assigned to 1000 subjects of knowledge, one-to-one, for the first draft of articles.

Active User Percentage (AUP): 0.4 (forty per cent) value is in parallel with findings about top ten Wikipedia editions (Ortega, Gonzalez-Barahona & Robles, 2008) in which authors discover that coordination and a considerable amount of work is supplied by the top ten per cent of members of the community, who both add new information and try to have a reputation in the circle of users. 0.2 and 0.6 values are used for variations.

General Interest Percentage (Administrator User Percentage – GIP): One per cent of active users (four in a thousand overall) are accepted as elite members who may actually reach to the administrator chair (users who are able to block others according to Wikipedia English edition) as well. However, in our context, administrator users are the ones that contribute the most, and prepare and edit Good Articles to make them conform to guidelines and regulations. This can be seen as a side effect of their substantial influence over the network or they can be thought as coordinators of the mass effort. This parameter has been varied to 0.005 and 0.015 values in the experiments.

Neighbourhood Dimension (ND): This parameter determines the area of interest for good samaritans and project leaders on the cartesian plane of information, and the neighbourhood dimension is the diameter of the Moore neighbourhood around the user node. Four units for the diameter of a neighbourhood means 9x9=81 unit square area around the user node is accessible for it, and on average eight editable articles are about the user's field of interest for base run. This parameter takes values 3 and 6 as the variation for observing influence over network characteristics.

Good Article Connection Count Threshold (GACCT): At least five edits are required for an article to become a good article candidate in base run. This value is the bare minimum because an article having four or fewer edits may lack valuable information or include considerable personal bias, as it may reflect only a few people's opinions. This parameter is altered to 2 and 8 values in the further experiments.

Good Article Multiplier (GAM): This value determines how much an article needs to be edited to be counted as a good article, as a multiple of the average degree of the article nodes in the network. The default value is 2.5, which means an article should have 25 or more edits before going into the good article window if the average edit count is 10. This parameter was first lowered to 1.5, but no considerable effect was observed in the results. This finding led to the notion that there is a threshold, and that greater values would affect the network. Therefore 5 and 7.5 values are used for variations, producing meaningful results.

Table 3 Variation of Input Parameters for the Experiments

	Lower Value	Base Run	Upper Value
Article Count	-	1000	-
User Count	-	1000	-
Active User Percentage (AUP)	0.2	0.4	0.6
General Interest Percentage (GIP)	0.005	0.01	0.015
Neighbourhood Dimension (ND)	3	4	6
Good Article Connection Count Threshold (GACCT)	2	5	8
Good Article Multiplier (GAM)	1.5	2.5	5, 7.5

CHAPTER 4

RESULTS

This section firstly represents average output metrics of resulting networks for variations of parameters in the experiments in Table 4 and Table 5. Succeeding sub-sections summarises and discusses the results according to sensitivity analysis of input parameters, and their effects on the specific network characteristics.

Table 4 Average Output Metrics of Networks for Base Run, Variations of Average User Percentage (AUP), General Interest Percentage (GIP), Neighbourhood Dimension (ND)

Avg. Output Metrics	Input Val.	Base Run	AUP= 0.2	AUP= 0.6	GIP= 0.005	GIP= 0.015	ND= 3	ND= 6
Total Edge Count	Affl. Net.	2987	1834	4201	2968	2923	1939	4225
	Art. Net.	12558	7265	21239	12783	11871	5392	25993
	Usr. Net.	5824	2300	10552	5417	5665	2916	9243
Small World Charc. Q	Usr. Net.	1.87	0.79	2.46	1.64	1.81	0.96	2.31
	Art. Net.	1.68	0.94	2.71	1.85	2.46	0.87	2.80
Path Length	Usr. Net.	4.58	8.75	3.32	5.02	4.87	8.62	3.60
	Art. Net.	4.76	8.41	3.23	4.77	3.32	9.48	3.09
Clustering Coefficient	Usr. Net.	0.62	0.50	0.59	0.60	0.64	0.60	0.60
	Art. Net.	0.58	0.58	0.64	0.64	0.59	0.60	0.63

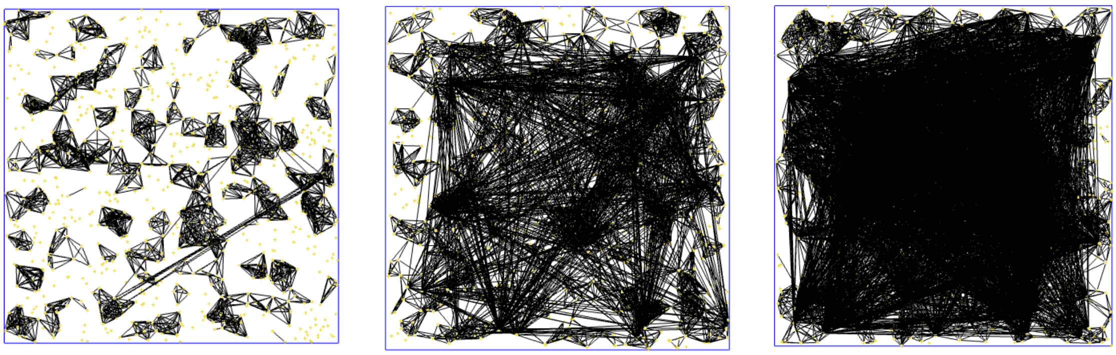
Table 5 Average Output Metrics of Networks for Base Run, Variations of Good Article Multiplier (GAM), Good Article Connection Count Threshold (GACCT)

Av. Output Metrics	Input Val.	Base Run	GAM= 1.5	GAM= 5	GAM= 7.5	GACCT= 2	GACCT= 8
Total Edge Count	Affl. Net.	2987	2996	2946	2885	3095	1834
	Art. Net.	12558	12717	11587	11136	19967	11228
	Usr. Net.	5824	5756	5408	5013	6230	4982
Small World Charc. Q	Usr. Net.	1.87	1.97	1.43	1.03	2.38	0.86
	Art. Net.	1.68	1.76	1.60	0.78	3.05	0.75
Path Length	Usr. Net.	4.58	4.62	5.84	9.18	3.40	9.99
	Art. Net.	4.76	4.68	5.56	10.71	2.99	10.71
Clustering Coefficient	Usr. Net.	0.62	0.66	0.61	0.68	0.59	0.62
	Art. Net.	0.58	0.60	0.65	0.61	0.66	0.58

Note: Average Total Edge Count values have been rounded to integer values in Tables 4 and 5.

Neighbourhood Dimension (ND)

Neighbourhood dimension (ND) is very effective in both increasing average edit count (edge count or degree) for articles, and reducing path length without a substantial loss in clustering coefficient, leading to a higher small world characteristic Q, suggesting a better structured, more informative encyclopaedia. When neighbourhood dimensions rise from 3 to 4 and then 6, average article and user degree grows more than proportionally, suggesting a very effective relationship. Another observation of particular interest here is that the article network does not stabilise after 50 ticks when neighbourhood dimension is 6. Even if the affiliation network and the user network's total degree count reaches a plateau and tends to stay constant, the article network's total degree count shows linear growth until the end of the experiment, when ND=6. When ND is 3, however, the main affiliation network and other networks get away from the being a proper "network". That is, they are sets of tiny disconnected components (in the size of the ND), and "one giant component to cover and connect most of the nodes" does not exist. These structural differences are traceable in Figure 13.



a. b. c.

Fig. 13 Article Network graph visualisations for experiments when neighbourhood dimension (ND) = 3 (a), ND = 4 (b) and ND = 6 (c).

Good Article Multiplier (GAM) and Good Article Connection Count Threshold (GACCT)

GAM and GACCT basically influence the system in the same fashion. GAM limits the good article selection process with a multiple of average degree of the network, whereas GACCT determines the lowest edit count for being included in the set. These two parameters have produced a significant result of this study. If we ease the selection process for the good articles (or the featured articles), the resulting interest rise improves the total quality of the encyclopaedia. For the base run, we used value 5 as the good article connection count threshold. If we raise the barrier, let's say GACCT=8, the affiliation network loses nearly half of its total edit count: 1834 (2987 for base run). If we would use more relaxed limits like GACCT=2, however, the article network total edge count almost doubles: 19967 (12257 for base run). Effects on the user network development may be seen in Figure 15 visually. We observe a similar pattern in path length and small-world characteristic metrics as well. User network path length and article path length duos are: (3.40,2.58) for GACCT=2, (4.58, 4.76) for base run, (9.99, 10.71) for GACCT=8. Small-world characteristic Q, likewise (2.38,3.05) for GACCT=2, (1.87, 1.68) for base run, and (0.86, 0.75) for GACCT=8. Similarly, these Q values for GAM are (1.87, 1.68) for GAM=2.5 (base run), (1.41,1.60) for GAM=5 and (1.03, 0.78) for GAM=7.5. Reduction of small-world characteristic of the network is traceable in Figure 14.

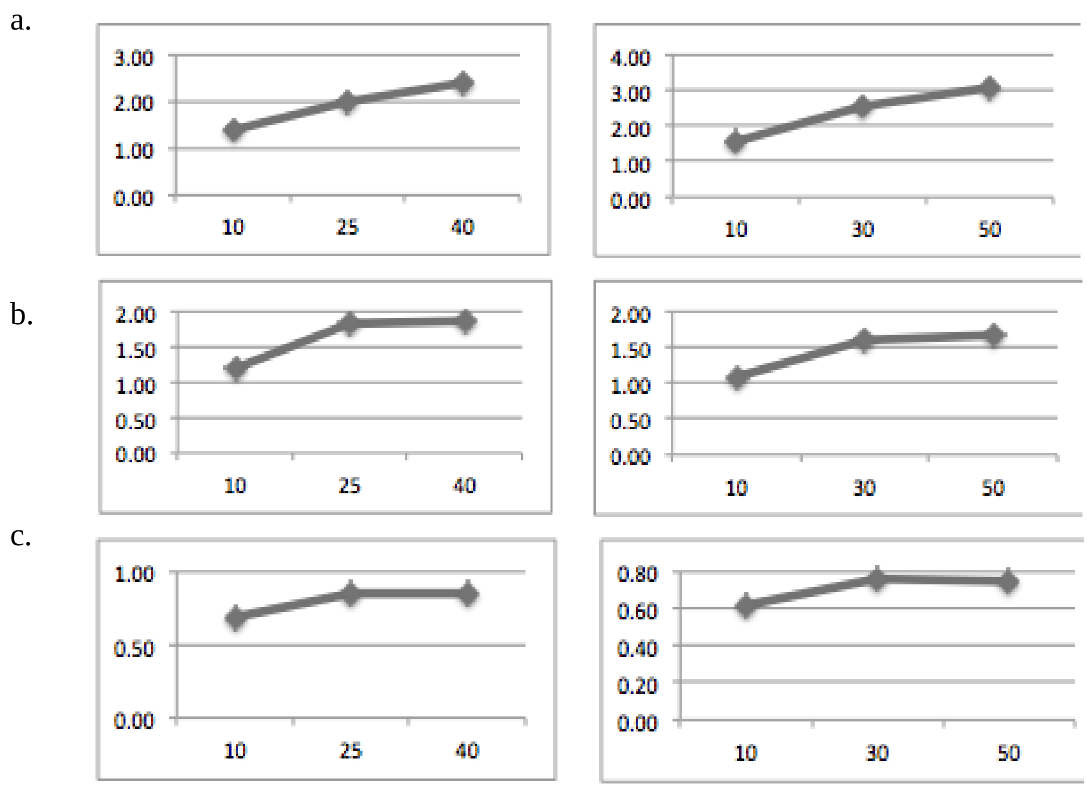
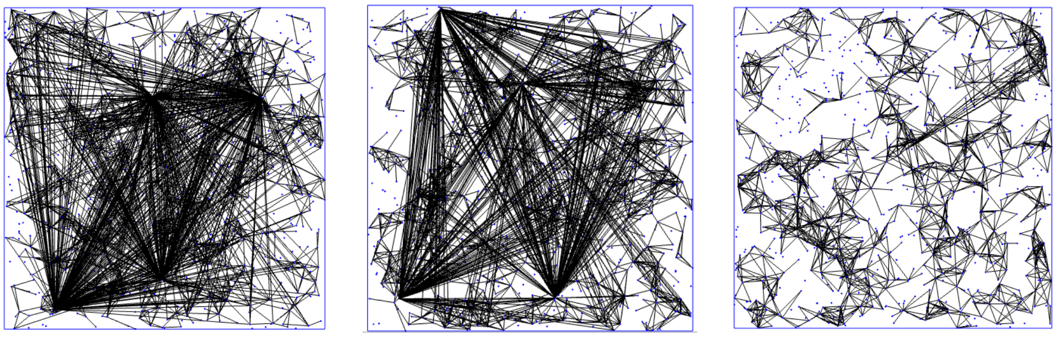


Fig. 14 Small-world characteristic Q development for user and article networks, affected by the variations of GACCT (GACCT = 2 in a, GACCT = 5 in b – base run, GACCT = 8 in c)

Additionally, we observe "divergent" good articles and zealots and reduced linearity (lower r^2) in lower barriers (low GACCT and GAM). These outputs clearly demonstrate reduced barriers for the elite set of articles (good or featured) not only populate and improve the set more, but also results in increased quality for the whole.



a. b. c.

Fig. 15 User Network graph visualisations for experiments when Good Article Connection Count Threshold (GACCT) = 2 (a), GACCT = 5 (b) and GACCT = 8 (c). These visualisations clearly demonstrate the “negative” impact of high standards for good article selection criteria on Wikipedia community network of our model.

Active User Percentage (AUP)

Active users (zealots) are the hard workers of the online community. If we somehow raise their relative population in the community, it is rather intuitive that overall quality improves. This notion is in parallel with our findings: it affects all three networks’ total edge counts with a very effective positive relationship and total edge counts increase asymmetrically. For instance, affiliation network includes 1834 edges when AUP=0.2, 2987 when AUP=0.4 (base run), and 4201 when AUP=0.6. Total edge development pattern for variations of AUP in Figure 16 supports this.

a.
b.
c.

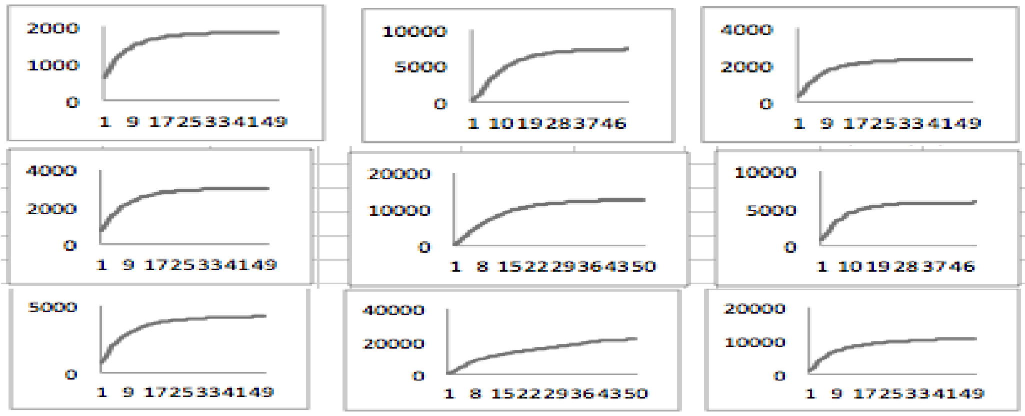


Fig. 16 Total edge development pattern for affiliation, user and article networks for AUP = 2 (a), AUP = 5 (b – base run) and AUP = 8 (c)

Same story holds true for path length in user network: 8.75 for AUP=0.2, 4.58 for base run, and 3.32 for AUP=0.6. Small-world characteristic Q is also affected considerably and interestingly. If AUP is too low (2) network does not even develop – the 10th, 25th, and 40th step Q values are (0.79, 0.73, 0.79), indicating that the network grows disconnected. More users and articles are divergent in degree correlation charts when percentage increases, pointing out more effort. Degree distribution charts point out same relationship: if the zealot percentage increases, the degree distribution shifts right for both article and user networks. Another observation here is that high AUP leads to higher differences of the experiment's degree distribution values, raising the standard deviation among the experiments.

<u>General Interest Percentage (GIP)</u>

General interest percentage determines ratio and population of administrators in the system. Basically, we failed to discover any substantial effect of GIP on the resulting output measures of the system. This is clearly observable in total edge counts of affiliation network, article network and user network values for changing GIP inputs: (2905.31, 12783.35, 5417.70) for GIP = 0.005; (2987.35, 12557.65, 5823.80) for GIP = 0.01 - base run; and (2923.50, 11871.35, 5665.80) for GIP = 0.015. The administrator count, as can be seen, not only does fail to dramatically increase edit count and quality in encyclopaedia, but also it may even slightly decrease it. Small-world characteristic Q values also support the ineffective factor status: user network Q values are, (1.64) for GIP=0.005, (1.87) for GIP=0.01; and (1.81) for GIP=0.015. GIP increases degree correlation between users and their neighbours as well (we have found out higher degree correlation to be an indicator of lesser quality for the encyclopaedia). There might be two distinct explanations for this obvious ineffectiveness of administrator counts in the dynamics of the model. Firstly, our model uses a seriously low percentage of general interest in our experiments: 0.005, 0.01, 0.015, therefore we might fail to assess the influence on the results. Secondly, it might stem from the very nature of the encyclopaedia we aim to simulate: It is not an administrator-driven encyclopaedia, and administrators are actually merely hardworking users. These explanations will be elaborated in the conclusion chapter.

Results for Specific Network Characteristics

Total Edge Count

Total edge count (degree) is synonym for sum of edit counts (user activity) in Affiliation Network, therefore total edge count of affiliation network is our primary quality metric for a newly developing encyclopaedia. Quality measuring by the count of editorial activity on an article is an established metric for Wikipedia in the literature as well (Wilkinson & Huberman, 2007) as edits are the indivisible units of effort for an article.

For our simulation, total edge count development tends to follow a general pattern – the chart of total edge counts against tick counts (time) has an asymptotic trend-line which eventually seems to get steady. This is because, first steps of the experiment are the most suitable period for expansion – project leaders' neighbourhoods are empty, and promoting to good article status is easy as average degree is low in the network. Eventually project leaders complete their collaboration efforts and good articles array freezes; leading the system to stabilise as shown in Figure 17 for base run (main experiment without variations in parameters).

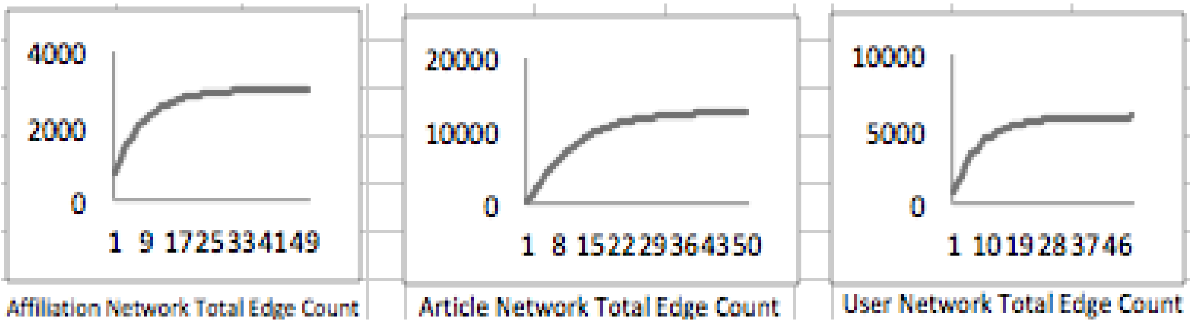


Fig. 17 Network total edge count development for base run.

This trend is observed all three types of networks (affiliation, user and article).

There are some exceptions for the stabilisation process. If the network cannot complete its development in 50 ticks and nodes still tend to connect at the end of the experiment, we continue to observe an increase trend in article network through the final steps of the experiment. This is a strong indicator for a better quality encyclopaedia for this parameter value. This is observable in experiments ND=6 and AUP=0.6 suggesting active user count and their area of interest are strong mediators of quality in the free encyclopaedia.

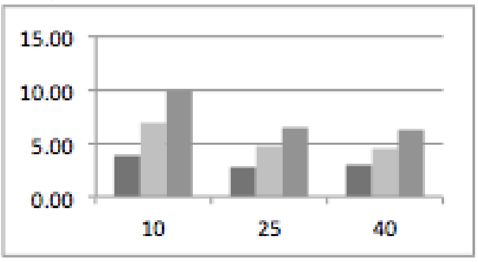
For base run, step 50 values for total edge counts are 2987.35 on average for affiliation network, 12557 for article network, and 5823 for user network. For comparison, other experiments' results are shown in Table 28 and Table 29.

Path Length

Path length value tends to decrease as the collaboration network develops and matures, as may be seen in Figure 18. Path length is found to be in an inverse relationship with neighbourhood dimension parameter and ND is very effective on the reduction of path length value. The relationship, however, cannot be defined as “correlation”, because path length value is more difficult to further reduce when the value is lower.

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.94	7.00	10.07
25	2.82	4.63	6.44
40	2.94	4.58	6.22

a.



Step	AVG-STDEV	AVG	AVG+STDEV
10	3.74	6.86	9.99
30	3.01	4.98	6.95
50	2.78	4.76	6.73

b.

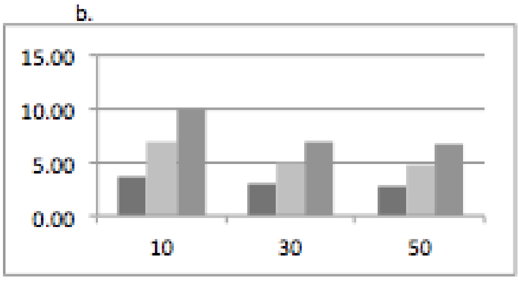


Fig. 18 Average path length development for user(a) and article(b) networks.

We may mention here about a certain “threshold” for path length value, and in our experiments this threshold seems to be close to 3. Of course the ultimate threshold for the path length of any network is 1, the value of fully connected network. Our experiments, on the other hand, tries to simulate a network of human community, of which several studies measured to have “six degrees of separation” (the path length is six) beginning with the famous experiment of Milgram (1967). Later studies found even lower path length values in the online world, for instance four degrees of separation for F*cebook.com, a social media network (Backstrom et al., 2012). This is parallel with our findings, as the lowest path length value we reached in the whole simulation is 3.09 (when ND = 6, Table 4).

Clustering Coefficient

Clustering Coefficient measurement gives unintuitive results, as it does not change and deviate much throughout our simulation, other than small-scale fluctuations of expected direction (for article network; path length = 0.63 when ND = 6 and path length = 0.60 when ND = 3; path length = 0.58 when active user percentage = 0.2 and path length = 0.64 when active user percentage = 0.6). This may be exemplified clustering coefficient development in base run in Figure 19.

Average Clustering Coefficient in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.25	0.61	0.98
25	0.30	0.62	0.94
40	0.29	0.62	0.95

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.14	0.55	0.95
30	0.18	0.58	0.98
50	0.18	0.58	0.98

a. b.

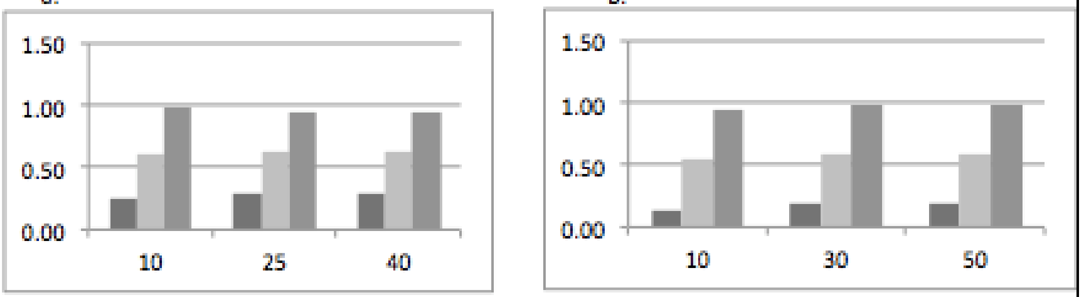


Fig. 19 Clustering coefficient development for user(a) and article(b) networks.

In parallel to these findings, Ingawale et al.'s paper (2009) demonstrated that clustering coefficient's change is relatively low in the development of Cebuano language Wikipedia. Additionally, the definition of small-world network includes having clustering coefficient similar to a regular lattice, with a considerably low path length (Watts & Strogatz, 1998), hinting a stable nature of clustering coefficient in the concept of graph theory. Clusters are closely linked to the concept of neighbourhoods (clustering coefficient is defined to be ratio of one node's connected neighbours to total count of neighbours), and our model is based on the neighbourhood concept (by neighbourhood connection algorithm), therefore existence of a stable clustering basis – at around sixty per cent – throughout the experiment, is not surprising.

Small-World Characteristic Q

Small-World Characteristic Q, the main measuring stick for quality and similarity-to-nature for networks, demonstrates a clear relationship with network maturity as seen in Figure 20, and with our input parameters. Neighbourhood

Average Small World Coefficient in User Network

Step	AVG
10	1.20
25	1.84
40	1.87

a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.09
30	1.61
50	1.68

b.

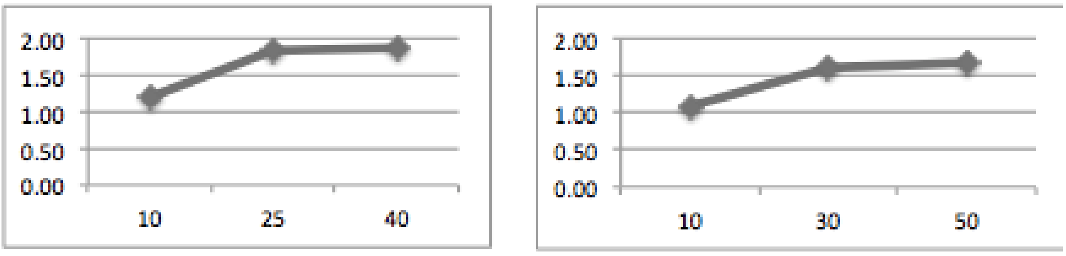


Fig. 20 Small world char. Q development for user(a) and article(b) networks.

dimension value affects small-world characteristic Q with a substantial positive relationship, that is, if ND value increases, the network “grows smaller” – average degree (average edit count and average quality) rises, while network becomes more tightly knitted together. This suggests, while neighbourhoods (clusters) still exist, they cover a larger area, gathering more information from users for articles. More members of these neighbourhoods are reached and edited by administrators, and they become up to the standards. In the big picture, we eventually have more standardised-and-high-quality articles, which means a higher quality encyclopaedia. This substantial relationship is shown to hold true for Active User Percentage (AUP,

Table 4). An inverse relationship between Q and good article parameters can be observed as well (GACCT and GAM, Table 5).

Degree Correlation

Collaboration network in our model holds a high degree of degree correlation as may be seen in Figure 21.

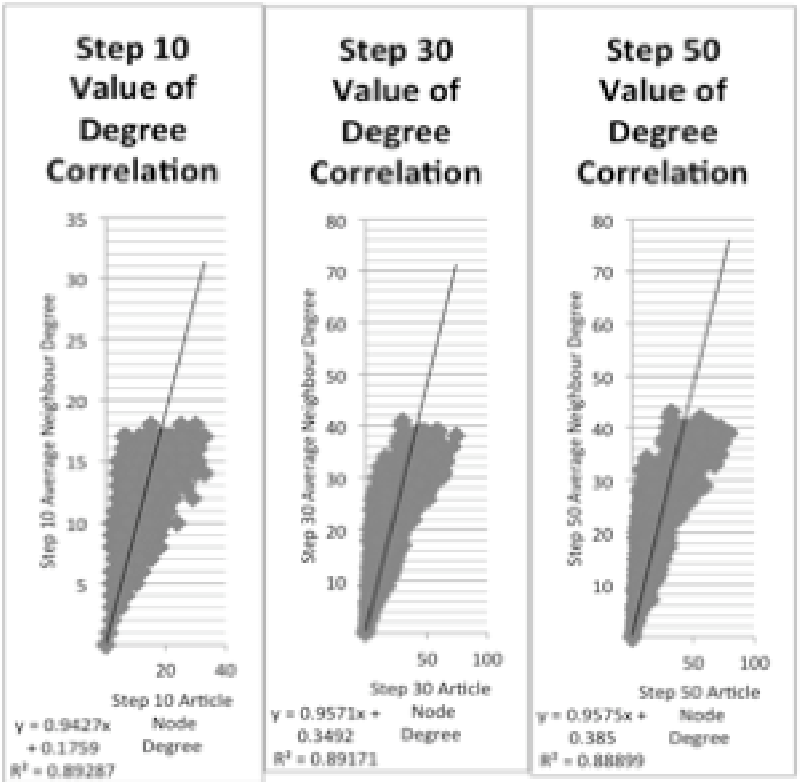


Fig. 21 Degree correlation development of article network for base run.

Degree correlation follows an interesting pattern when network qualifications alter. It seems to in an inverse relationship with the quality of emerging encyclopaedia in our simulation experiments. When neighbourhood dimension (ND), good article multiplier (GAM) or active user percentage (AUC) value enhances zealot activity, and this leads to more edits for the average article and increases quality. Throughout this effect, however, zealots (active users) and good articles positively “diverge” from normal nodes in their network – they seem to have a different pattern of collaboration network, and they therefore do not abide by the linear degree correlation structure of good samaritans. Higher value nodes of article network “bend” the upper part of degree correlation line, reducing linearity, therefore decreasing r^2 value of linear function used to express the relationship.

Degree Distribution

Degree distribution observed in the experiments have a specific structure for article and user networks. User networks’ degree distribution generally represents an asymptotic decrease with high population of users with few edits representing good samaritans. However users’ degree distribution chart includes an interesting “bump” in the tail section as well and this bump points out zealots’ average edit count. This way users’ degree distribution may be considered as a combination of two distributions – one for good samaritans and one for zealots. This common pattern is observable in Figure 22.

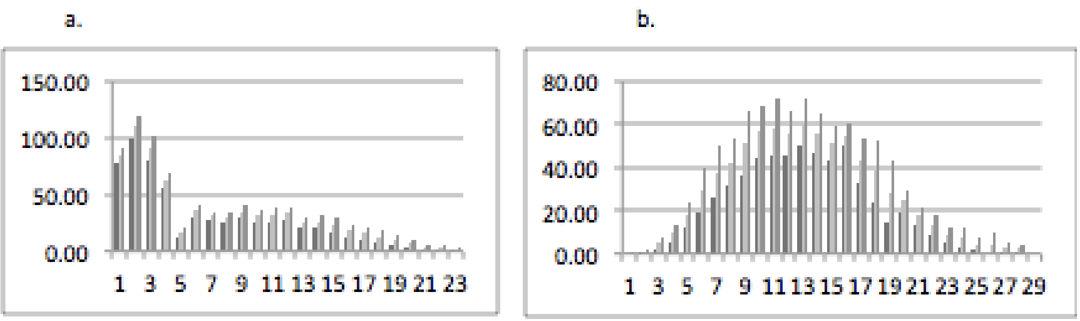


Fig. 22 Degree distribution in user(a) and article(b) network s for base run.

Effects of two input parameters are clearly observable on user degree distribution: Active user percentage (AUP) and neighbour dimension (ND). As they increase, they “thicken” and stretch the tail of the distribution suggesting main effects being on the zealots’ contributions. Numbers express similar results: For AUP, sum of average count of users who made more edits like 17-19 is 1.95 for AUP=0.2, 39.15 for AUP=0.4 and 95.95 for AUP=0.6; for ND we observe a shift of local extremum of the ‘bump’ to the right, the extremum point is 6 degrees for ND=3, 9 degrees for ND=4 and 19 degrees for ND=6. Good Article Connection Count Threshold (GACCT) and Good Article Multiplier (GAM), on the other hand, make the tail thinner and weaken the activity of zealots considerably. The average count of users having between 14 and 16 edits (zealots) is 79.10 when GACCT=2, 67.35 when GACCT=5 and 58.45 when GACCT=8; whereas those having between 18 and 20 edits are 30.15 when GAM=2.5, 18.55 when GAM=5, and 12.30 when GAM=7.5.

CHAPTER 5

CONCLUSION

This study has examined and discussed the application of agent-based modelling and graph theory concepts on the development of an online community for collaborative content creation. As the outcome of graph-theoretical agent-based simulation experiments, we reached significant findings that support these three results:

1. A rise in the relative population of active users' (project leaders and administrators) leads to more effort for the information quality and results in more comprehensive and informative encyclopaedia.
2. Enlargement and diversity of these active users' area of interest (neighbourhood dimension) make the resulting network more tightly knitted and improve overall quality of the encyclopaedia.
3. Provided that the elite set of articles (good articles and featured articles) are allowed to grow and the selection process is easy (good article multiplier and good article connection count threshold are low), information quality and average edit count increase not only for these articles, but also for the whole encyclopaedia, leading to better quality.

These findings clearly deserve a deeper analysis and possible use for a better Wikipedia.

That higher relative population and percentage of zealots increase information content of the encyclopaedia is obviously intuitive. However, our simulation demonstrates that this increase is not a mere pile-up of unstructured information. More zealots additionally mean that a better organised network, and more structured well-represented information. Collaboration network becomes more tightly connected with members are linked with less degrees of separation (shorter path length). How to achieve a higher rate of repeatedly – contributing active users, is an essential question, therefore. Wikipedia policymakers are aware of this need as well. They attempt to encourage anonymous users to have an online identity and account – the compulsory starting point for a regularly contributing member. Wikipedia has special pages dedicated to persuade its contributors to sign up for an online account ("Wikipedia: Why create an account?", n.d.). We propose an extension for this apparently required act. If the user is an anonymous one, Wikipedia stores their contribution using their IP – both for statistical and security reasons. We propose to send a cookie to anonymous users' computers and use it to understand if the user has contributed before. If they did, the next time they open a Wikipedia webpage, we may show a banner recognizing their previous contributions, and kindly requesting more. This way, Wikipedia might benefit more from its anonymous users and we might reach a novel concept: an anonymous zealot. Because many users deliberately stay anonymous and their contributions are the most constructive (Antony, Smith & Williamson, 2007), their repeated efforts might really help make the free encyclopaedia a better place.

Diversifying good samaritans' and project leaders' topics of interest to spread their efforts (neighbourhood dimension concept) is another method we demonstrated to positively affect the health of the community network and quality of resulting encyclopaedia. The persuasion of experts on a subject to contribute on related subjects should therefore be another target. In order to achieve this, we propose an article-oriented mechanism. This is because, in Wikipedia's chaotic and anonymous environment, not only would persuasion of users be nearly impossible, but also even identifying “which user is an expert on what” information would be too difficult. Rather, we recommend an increase of the concentration of internal links in the high-quality articles’ text. Links are blue-coloured words in the text, which take the reader to further information about the subject. The idea we propose is as follows: if we utilise a number of bots (computer scripts of Wikipedia) to increase link concentration in good and featured articles, as many expert users have worked on these articles to raise them to elite status, we encourage these experts to contribute on those linked articles as well. This is because red links (articles which do not exist) and blue links connecting to stubs, will get the attention of the expert group who enriched the article at hand.

As the last and probably the most useful method we propose is to ease the selection process for, and to extend the set of, good and featured articles. Our model of collaboration network of Wikipedia and succeeding simulation indicate that higher population of this elite set results in the improvement of quality for the whole product. Joining to the set of good and featured articles leads to a surge in interest for the new members of the set, and the higher quantity of the set enables the encyclopaedia to benefit more from this surge of interest. We propose a novel approach for good/featured article selection process, by automising it. In various studies, word count (Blumenstock, 2008) and edit count (Wilkinson & Huberman, 2007) have been shown to be direct indicators of quality for articles. We propose two additional metrics: citation count and bringing back “article score” by users (score section is now removed from Wikipedia). These four measurements would probably provide a high success in differentiating higher and normal quality articles. The rise of interest in these novel good/featured articles would result in more effort done for them, benefitting every article of the free encyclopaedia. Additionally, this approach would give readers an opportunity to utilise more information from Wikipedia. “Currently, out of the 4,600,461 articles on Wikipedia, 20,511 are categorized as good articles (about 1 in 225) and 4,364 are listed as featured articles (about 1 in 1,060)” (Wikipedia: Good articles, 2014), and the scarcity of these status tags leaves users no indicator of quality for most articles. Therefore, a general measurement stick would be of help for users to decide if they should use the information or not. In Wikia (the next for-profit project of one of Wikipedia founders, Jimmy Wales), the WAM Score is utilised for measuring the quality of Wikis, for example (“How is the WAM calculated?”, 2014). Additionally, this method may allow for good/normal separation in every academic field for Wikipedia, and may be used for creation of a nearly-complete encyclopaedia consisting of only high quality articles (e.g. for offline use). Therefore, advantages of this four-metric automatic assessment of articles would be multi-dimensional.

Notable constraints included bounded environment for collaboration network, constant amount of user and article nodes throughout the simulation, exclusion of vandalism behaviour. Future work on the subject might prefer a growing structure for online community as well as an expanding design for the cartesian plane of information. One of the main problematic areas for the free encyclopaedia, the vandalist behaviours of users, ought to be included in subsequent work on the subject. This could be done by randomly omitting the edges in collaboration network, representing deletion of a user’s efforts. Systematic vandalism and edit wars might be worked on as well.

Our study may find an important application area in Knowledge Management Systems (KMS) for corporations. KMS software is a basis for employee-collaboration for creation of training documents for new members. A similar or specialised model might be developed for examination of such systems as well. A KMS network would include directors and coordinators (administrators), expert, seasoned employees (project leaders) and one-shot contributors (good samaritans) as well. Such an approach would be of assistance for better management of KMS environments.

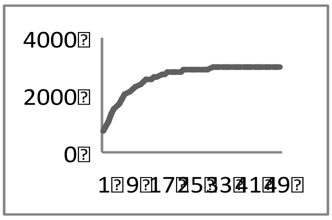
Lastly, we want to speculate on the basis, the essence, the main driving force of Wikipedia. Throughout the research and modelling process, Wikipedia never stopped astonishing us. Its quality, lack of formal rules and hierarchy, and endless journey always for the better, more than amazed us many times. We felt that – although this is open to debate and further research – the main driving force of Wikipedia is "goodness". It is an anti-thesis of the daily life consisting of a constant struggle for interests, and conflict of interests. At least for good samaritans, or dedicated editors (including the one who made over one million edits (Waugh, 2012)) however, there is no observable interest in giving information to the world. Therefore, what we understand here is, we, as human beings, always have a "good", or altruistic side. However, we cannot reflect this side of us in normal daily life as it might be abused or perceived as powerlessness. Wikipedia, however, provides us with the perfect place to reflect this benevolent side. We spread our information to the world, without introducing ourselves, for the greater good. This is not explicable with anything other than altruism. This is why Wikipedia might be taken as a reason to believe in the benevolent side of the human nature…

APPENDIX: OUTPUT DATA

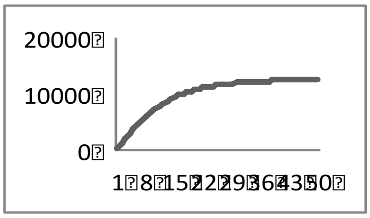
Base Run, n=20 Replications Results

Table 6 Total Edge Count Development for Base Run

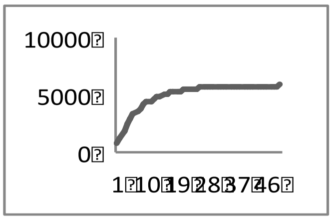
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	689.12	699.40	709.68	285.61	291.30	296.99	656.26	687.70	719.14
Step 2	1026.55	1043.55	1060.55	1012.84	1036.40	1059.96	1219.59	1278.60	1337.61
Step 3	1290.32	1310.60	1330.88	1927.18	1961.65	1996.12	1784.88	1875.45	1966.02
Step 4	1503.27	1530.50	1557.73	2868.24	2929.30	2990.36	2298.25	2416.50	2534.75
Step 5	1690.16	1720.05	1749.94	3779.28	3866.65	3954.02	2770.42	2901.25	3032.08
Step 6	1838.32	1878.90	1919.48	4583.27	4724.05	4864.83	3141.58	3291.55	3441.52
Step 7	1973.20	2015.30	2057.40	5363.01	5511.10	5659.19	3453.31	3616.50	3779.69
Step 8	2082.87	2129.40	2175.93	6008.83	6194.75	6380.67	3716.84	3890.30	4063.76
Step 9	2183.79	2229.95	2276.11	6621.20	6810.10	6999.00	3957.34	4130.00	4302.66
Step 10	2270.01	2316.45	2362.89	7194.30	7357.00	7519.70	4146.29	4323.80	4501.31
Step 11	2348.39	2396.50	2444.61	7694.84	7867.35	8039.86	4324.41	4501.80	4679.19
Step 12	2415.62	2463.20	2510.78	8159.10	8315.05	8471.00	4464.43	4652.40	4840.37
Step 13	2476.45	2525.05	2573.65	8565.59	8738.05	8910.51	4596.13	4789.30	4982.47
Step 14	2526.96	2576.85	2626.74	8889.64	9093.35	9297.06	4706.10	4903.60	5101.10
Step 15	2568.90	2622.55	2676.20	9184.81	9420.40	9655.99	4801.35	5011.00	5220.65
Step 16	2609.12	2663.20	2717.28	9490.19	9729.60	9969.01	4889.94	5107.10	5324.26
Step 17	2642.50	2698.20	2753.90	9751.08	10002.30	10253.52	4965.38	5188.30	5411.22
Step 18	2673.25	2730.10	2786.95	9980.56	10246.55	10512.54	5026.43	5257.20	5487.97
Step 19	2700.20	2758.00	2815.80	10189.54	10477.05	10764.56	5096.62	5325.40	5554.18
Step 20	2726.63	2783.50	2840.37	10397.06	10688.55	10980.04	5150.40	5379.80	5609.20
Step 21	2749.49	2807.75	2866.01	10592.41	10891.25	11190.09	5197.79	5436.30	5674.81
Step 22	2767.48	2828.05	2888.62	10746.59	11081.55	11416.51	5244.46	5492.00	5739.54
Step 23	2782.28	2844.75	2907.22	10872.23	11227.85	11583.47	5280.19	5529.50	5778.81
Step 24	2796.09	2859.75	2923.41	10985.16	11360.00	11734.84	5310.00	5560.10	5810.20
Step 25	2809.29	2874.85	2940.41	11099.47	11486.00	11872.53	5340.78	5590.90	5841.02
Step 26	2820.30	2885.90	2951.50	11187.65	11584.95	11982.25	5360.08	5616.80	5873.52
Step 27	2828.75	2897.00	2965.25	11262.81	11683.05	12103.29	5374.55	5637.90	5901.25
Step 28	2837.41	2907.05	2976.69	11338.22	11775.55	12212.88	5393.06	5658.80	5924.54
Step 29	2846.25	2916.40	2986.55	11429.05	11867.55	12306.05	5412.27	5679.90	5947.53
Step 30	2855.27	2924.00	2992.73	11502.24	11938.20	12374.16	5429.60	5693.70	5957.80
Step 31	2861.84	2931.05	3000.26	11557.12	12007.15	12457.18	5443.62	5708.70	5973.78
Step 32	2867.24	2937.35	3007.46	11603.06	12069.75	12536.44	5453.73	5721.40	5989.07
Step 33	2872.61	2943.20	3013.79	11643.71	12120.60	12597.49	5462.12	5732.20	6002.28
Step 34	2877.16	2947.90	3018.64	11684.30	12165.80	12647.30	5467.77	5740.20	6012.63
Step 35	2880.42	2952.45	3024.48	11718.69	12213.80	12708.91	5479.00	5751.80	6024.60
Step 36	2882.70	2955.85	3029.00	11739.44	12254.30	12769.16	5484.65	5760.20	6035.75
Step 37	2886.09	2959.75	3033.41	11770.94	12290.50	12810.06	5490.25	5767.90	6045.55
Step 38	2888.15	2962.65	3037.15	11785.58	12311.90	12838.22	5494.43	5773.20	6051.97
Step 39	2890.83	2965.75	3040.67	11813.50	12341.90	12870.30	5499.65	5778.70	6057.75
Step 40	2893.42	2968.65	3043.88	11836.12	12370.10	12904.08	5505.13	5784.90	6064.67
Step 41	2896.10	2971.80	3047.50	11865.22	12402.15	12939.08	5511.81	5790.70	6069.59
Step 42	2897.89	2973.90	3049.91	11878.38	12415.95	12953.52	5516.30	5794.90	6073.50
Step 43	2900.85	2976.95	3053.05	11908.28	12447.10	12985.92	5522.82	5800.90	6078.98
Step 44	2903.04	2979.10	3055.16	11934.07	12472.85	13011.63	5530.54	5806.20	6081.86
Step 45	2904.15	2980.50	3056.85	11948.00	12486.10	13024.20	5533.03	5809.90	6086.77
Step 46	2905.01	2982.00	3058.99	11958.34	12502.75	13047.16	5534.62	5812.80	6090.98
Step 47	2906.50	2983.85	3061.20	11971.31	12523.85	13076.39	5536.05	5815.80	6095.55
Step 48	2907.40	2985.00	3062.60	11980.64	12532.50	13084.36	5538.00	5818.20	6098.40
Step 49	2908.43	2985.90	3063.37	11986.38	12539.45	13092.52	5541.48	5820.10	6098.72
Step 50	2909.81	2987.35	3064.89	12004.82	12557.65	13110.48	5544.20	5823.80	6103.40



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

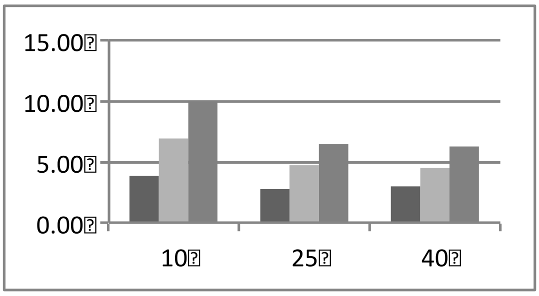
Fig. 23 Network total edge count development for base run.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.94	7.00	10.07
25	2.82	4.63	6.44
40	2.94	4.58	6.22

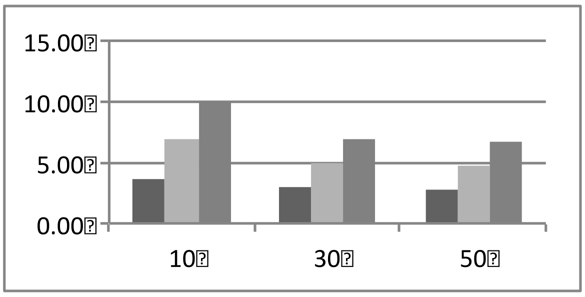
a.



Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.74	6.86	9.99
30	3.01	4.98	6.95
50	2.78	4.76	6.73

b.

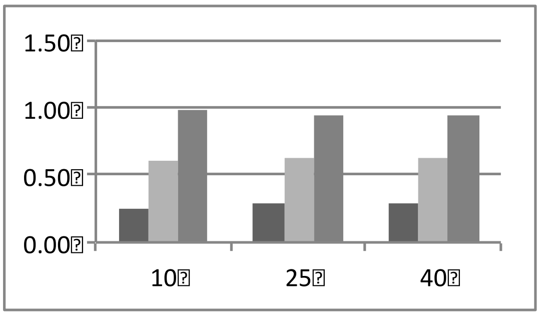


a) Pathlength development for user(a) and article(b) networks for Base Run.

Average Clustering Coefficient in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.25	0.61	0.98
25	0.30	0.62	0.94
40	0.29	0.62	0.95

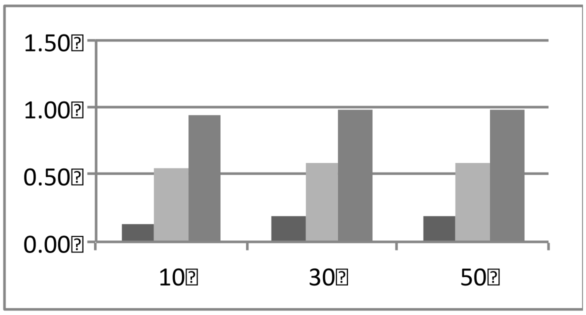
a.



Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.14	0.55	0.95
30	0.18	0.58	0.98
50	0.18	0.58	0.98

b.

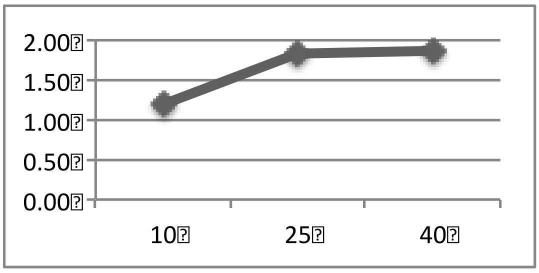


b) Clustering coefficient development for user(a) and article(b) networks for Base Run.

Average Small World Coefficient in User Network

Step	AVG
10	1.20
25	1.84
40	1.87

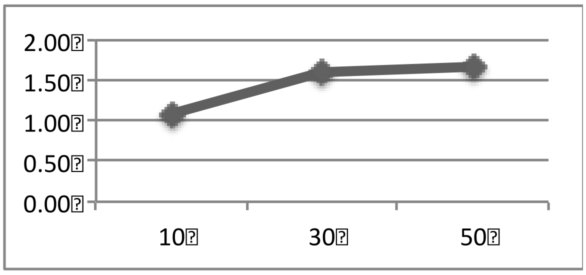
a.



Average Small World Coefficient in Article Network

Step	AVG
10	1.09
30	1.61
50	1.68

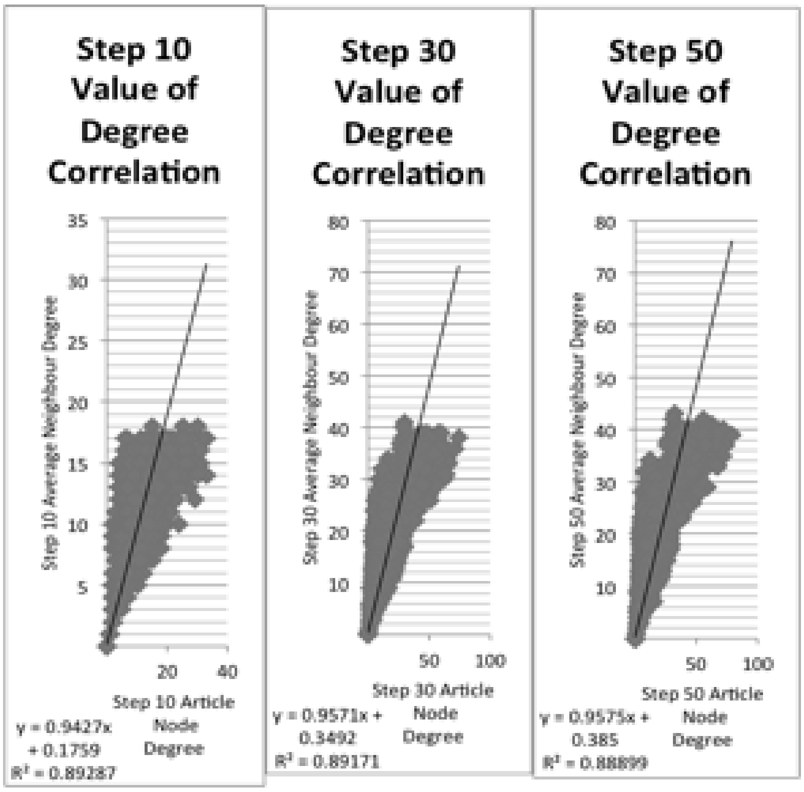
b.



c) Small World coefficient development for user(a) and article(b) networks for Base Run

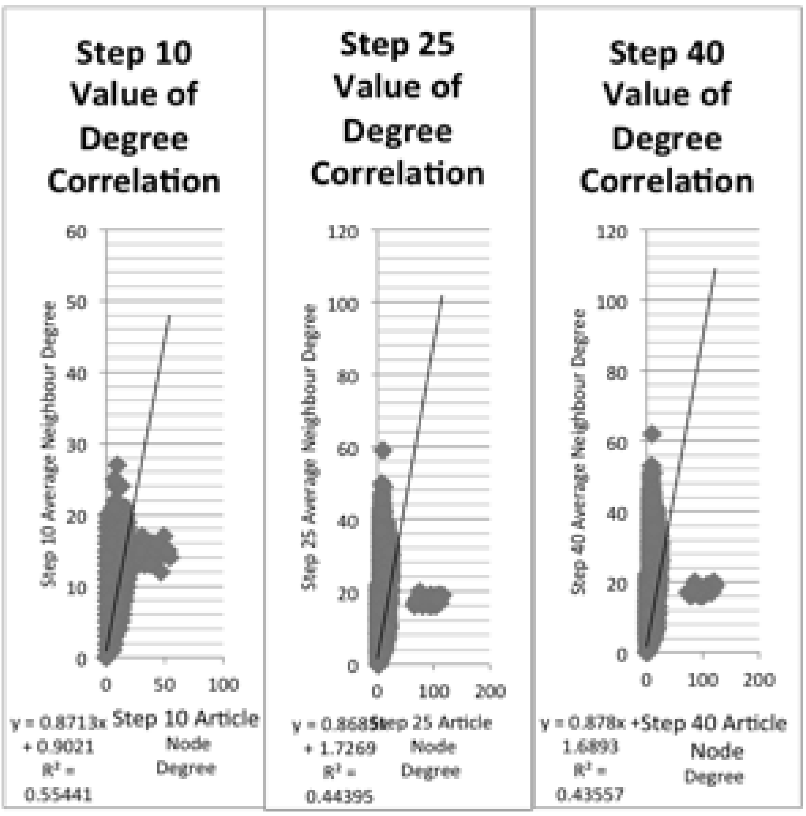
Fig. 24 Path length (a), clust. coeff. (b), small world char. Q (c) development for base run.

Article Network Degree Correlation Values



a) Degree correlation development for article network for Base Run.

User Network Degree Correlation Values



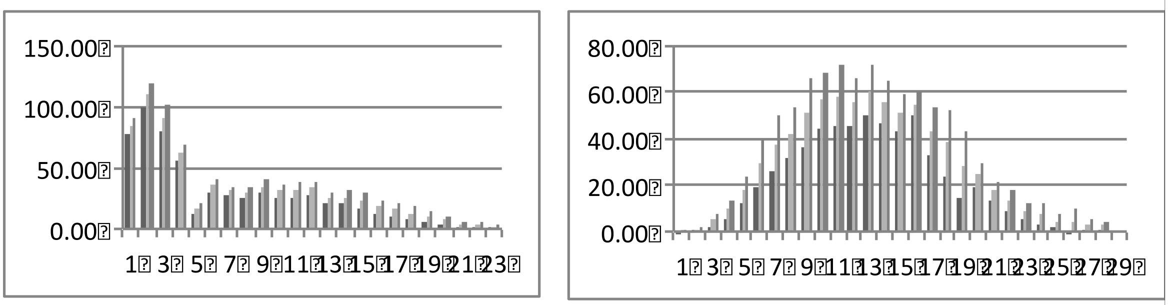
b) Degree correlation development for user network for Base Run

Fig. 25 Degree correlation development for base run

Table 7 Article and User Network Degree Distribution for Base Run.

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG+STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG+STDEV
1	78.30	84.25	90.20	1	-0.35	0.25	0.85
2	99.44	109.90	120.36	2	0.02	1.17	2.31
3	81.13	91.10	101.07	3	2.46	4.83	7.21
4	54.98	62.25	69.52	4	5.01	9.42	13.82
5	11.60	16.75	21.90	5	11.89	17.58	23.28
6	30.89	36.30	41.71	6	18.90	29.25	39.60
7	28.41	31.70	34.99	7	26.16	38.00	49.84
8	24.81	29.75	34.69	8	31.40	42.58	53.77
9	28.97	34.40	39.83	9	36.88	51.83	66.79
10	25.22	30.95	36.68	10	44.94	56.75	68.56
11	25.95	32.70	39.45	11	45.06	58.50	71.94
12	26.97	33.15	39.33	12	45.42	55.92	66.42
13	21.33	25.85	30.37	13	49.67	60.58	71.50
14	20.32	26.05	31.78	14	47.20	56.25	65.30
15	17.43	23.30	29.17	15	43.57	51.58	59.60
16	13.20	18.20	23.20	16	49.75	54.92	60.09
17	11.03	15.75	20.47	17	32.95	43.17	53.38
18	8.06	12.95	17.84	18	24.26	38.17	52.08
19	6.44	10.45	14.46	19	14.00	28.67	43.34
20	4.56	7.15	9.74	20	19.47	24.75	30.03
21	1.45	3.95	6.45	21	13.85	17.75	21.65
22	1.64	3.60	5.56	22	9.18	13.83	18.49
23	0.12	1.40	2.68	23	5.61	9.08	12.56
				24	3.54	7.67	11.80
				25	1.52	4.67	7.81
				26	-0.26	4.58	9.42
				27	0.90	2.92	4.94
				28	1.02	2.83	4.65

a. b.



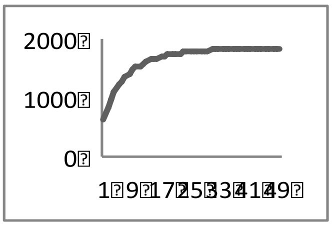
Degree distribution of User(a) and Article(b) networks.

Fig. 26 Degree distribution for base run.

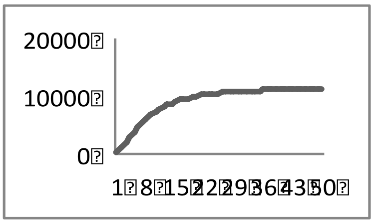
Good Article Connection Count Threshold (GACCT) = 8, n=20 Rep. Results

Table 8 Total Edge Count Development for GACCT = 8

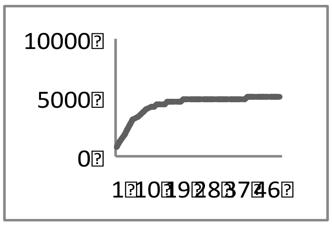
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	614.94	623.75	632.56	279.99	287.25	294.51	654.12	681.90	709.68
Step 2	824.97	837.85	850.73	1000.69	1023.75	1046.81	1195.75	1238.70	1281.65
Step 3	969.58	986.05	1002.52	1901.74	1950.85	1999.96	1756.29	1814.10	1871.91
Step 4	1088.61	1107.85	1127.09	2811.95	2890.50	2969.05	2241.22	2312.70	2384.18
Step 5	1186.83	1207.40	1227.97	3679.32	3791.15	3902.98	2647.29	2721.50	2795.71
Step 6	1264.56	1289.55	1314.54	4513.71	4659.10	4804.49	3004.66	3080.20	3155.74
Step 7	1331.92	1358.80	1385.68	5262.15	5435.95	5609.75	3286.80	3373.70	3460.60
Step 8	1390.19	1418.55	1446.91	5895.93	6125.05	6354.17	3513.83	3612.35	3710.87
Step 9	1441.27	1470.25	1499.23	6466.30	6724.00	6981.70	3695.07	3804.75	3914.43
Step 10	1482.22	1513.95	1545.68	6948.47	7253.60	7558.73	3851.88	3969.25	4086.62
Step 11	1517.46	1549.60	1581.74	7388.15	7713.75	8039.35	3987.43	4107.65	4227.87
Step 12	1548.64	1581.45	1614.26	7790.56	8120.00	8449.44	4099.84	4216.00	4332.16
Step 13	1577.77	1610.55	1643.33	8136.47	8478.35	8820.23	4198.75	4321.50	4444.25
Step 14	1603.21	1636.35	1669.49	8460.28	8803.05	9145.82	4281.45	4407.35	4533.25
Step 15	1624.94	1659.10	1693.26	8700.83	9073.95	9447.07	4357.39	4482.65	4607.91
Step 16	1641.84	1678.45	1715.06	8929.28	9306.65	9684.02	4418.82	4546.35	4673.88
Step 17	1658.11	1695.45	1732.79	9123.44	9509.45	9895.46	4462.21	4596.70	4731.19
Step 18	1672.94	1710.70	1748.46	9299.52	9689.85	10080.18	4506.65	4640.75	4774.85
Step 19	1686.39	1724.75	1763.11	9455.43	9866.70	10277.97	4545.12	4683.00	4820.88
Step 20	1696.91	1737.05	1777.19	9611.19	10018.20	10425.21	4584.62	4721.70	4858.78
Step 21	1707.46	1748.00	1788.54	9743.91	10164.15	10584.39	4621.81	4756.30	4890.79
Step 22	1716.00	1756.90	1797.80	9852.16	10276.00	10699.84	4646.44	4784.20	4921.96
Step 23	1724.10	1765.60	1807.10	9945.68	10373.05	10800.42	4669.29	4808.70	4948.11
Step 24	1732.10	1773.30	1814.50	10017.34	10452.30	10887.26	4690.23	4828.05	4965.87
Step 25	1737.90	1780.70	1823.50	10087.08	10529.30	10971.52	4709.70	4845.95	4982.20
Step 26	1744.12	1786.70	1829.28	10146.17	10595.70	11045.23	4722.58	4861.05	4999.52
Step 27	1749.64	1791.95	1834.26	10201.11	10655.10	11109.09	4737.23	4874.35	5011.47
Step 28	1754.48	1796.85	1839.22	10256.27	10713.70	11171.13	4747.17	4887.95	5028.73
Step 29	1757.94	1800.60	1843.26	10324.02	10776.70	11229.38	4759.80	4900.45	5041.10
Step 30	1762.34	1805.20	1848.06	10368.49	10824.45	11280.41	4768.89	4909.25	5049.61
Step 31	1765.84	1808.95	1852.06	10399.58	10868.80	11338.02	4775.01	4919.05	5063.09
Step 32	1769.24	1812.45	1855.66	10434.35	10905.25	11376.15	4784.73	4927.00	5069.27
Step 33	1771.88	1815.35	1858.82	10473.10	10940.90	11408.70	4790.26	4934.50	5078.74
Step 34	1774.40	1817.75	1861.10	10508.00	10979.90	11451.80	4797.46	4943.15	5088.84
Step 35	1776.94	1819.60	1862.26	10528.91	11011.55	11494.19	4802.58	4948.85	5095.12
Step 36	1778.25	1821.30	1864.35	10552.83	11043.25	11533.67	4805.79	4953.15	5100.51
Step 37	1779.76	1823.25	1866.74	10574.44	11068.90	11563.36	4808.32	4956.65	5104.98
Step 38	1780.90	1824.50	1868.10	10595.74	11089.50	11583.26	4810.85	4959.75	5108.65
Step 39	1782.36	1826.00	1869.64	10610.29	11107.75	11605.21	4813.09	4962.35	5111.61
Step 40	1783.56	1827.20	1870.84	10622.65	11122.50	11622.35	4814.26	4964.15	5114.04
Step 41	1784.62	1828.40	1872.18	10628.90	11138.15	11647.40	4816.79	4966.65	5116.51
Step 42	1785.25	1829.10	1872.95	10649.00	11156.00	11663.00	4819.48	4969.25	5119.02
Step 43	1785.52	1829.85	1874.18	10663.16	11167.25	11671.34	4821.07	4971.65	5122.23
Step 44	1786.56	1830.65	1874.74	10676.25	11178.65	11681.05	4823.27	4973.45	5123.63
Step 45	1787.31	1831.50	1875.69	10680.24	11184.75	11689.26	4824.32	4974.65	5124.98
Step 46	1788.10	1832.10	1876.10	10688.71	11194.60	11700.49	4825.11	4975.85	5126.59
Step 47	1788.54	1832.90	1877.26	10693.93	11202.60	11711.27	4825.64	4978.40	5131.16
Step 48	1788.91	1833.30	1877.69	10698.76	11215.45	11732.14	4827.40	4979.70	5132.00
Step 49	1789.65	1834.00	1878.35	10707.42	11221.25	11735.08	4827.93	4980.40	5132.87
Step 50	1789.97	1834.75	1879.53	10711.37	11228.70	11746.03	4829.16	4982.00	5134.84



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

Fig. 27 Network total edge count development for GACCT = 8.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

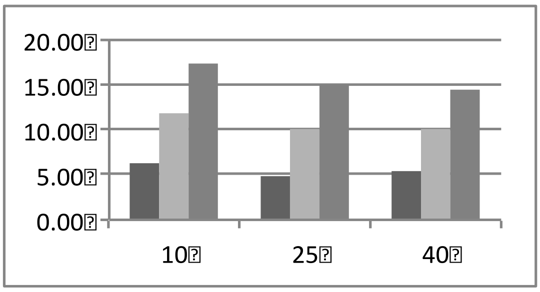
Step	AVG-STDEV	AVG	AVG+STDEV
10	6.28	11.76	17.24
25	4.88	9.97	15.06
40	5.50	9.99	14.48

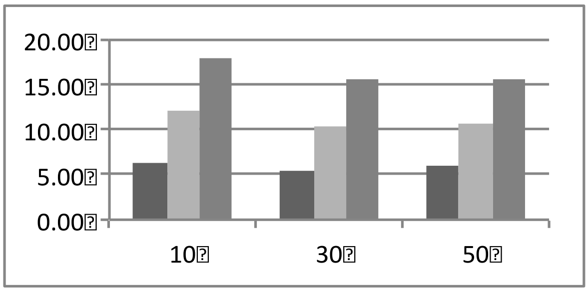
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	6.39	12.21	18.03
30	5.49	10.51	15.53
50	5.89	10.71	15.53

b.





a) Pathlength development for user(a) and article(b) networks when GACCT = 8

Average Clustering Coefficient in User Network

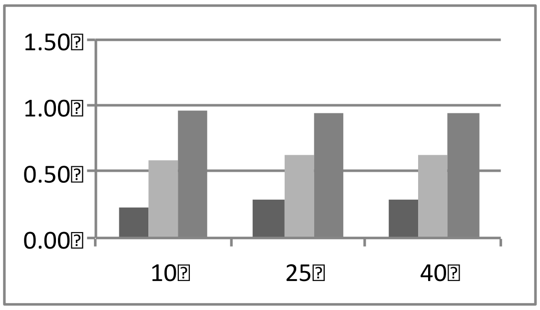
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.22	0.59	0.96
25	0.29	0.62	0.95
40	0.30	0.62	0.95

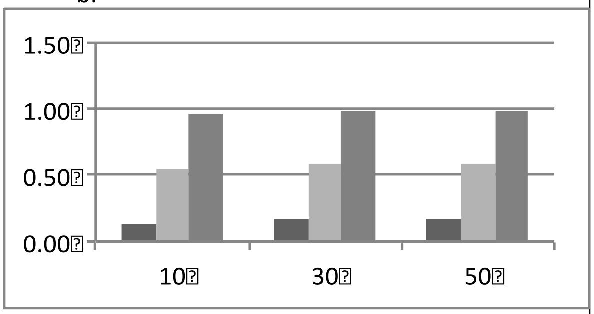
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.14	0.55	0.96
30	0.18	0.58	0.98
50	0.18	0.58	0.99

b.





b) Clustering coefficient development for user(a) and article(b) networks when GACCT = 8

Average Small World Coefficient in User Network

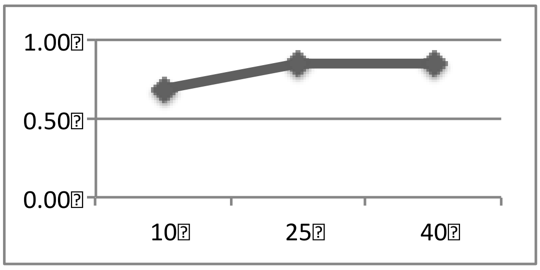
Step	AVG
10	0.69
25	0.86
40	0.86

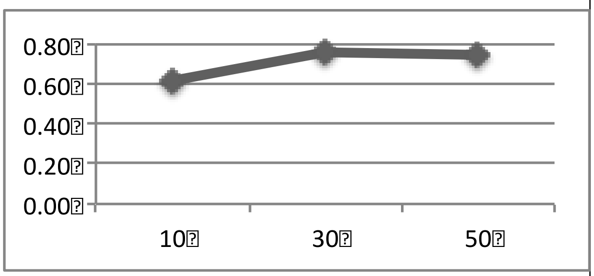
a.

Average Small World Coefficient in Article Network

Step	AVG
10	0.62
30	0.76
50	0.75

b.

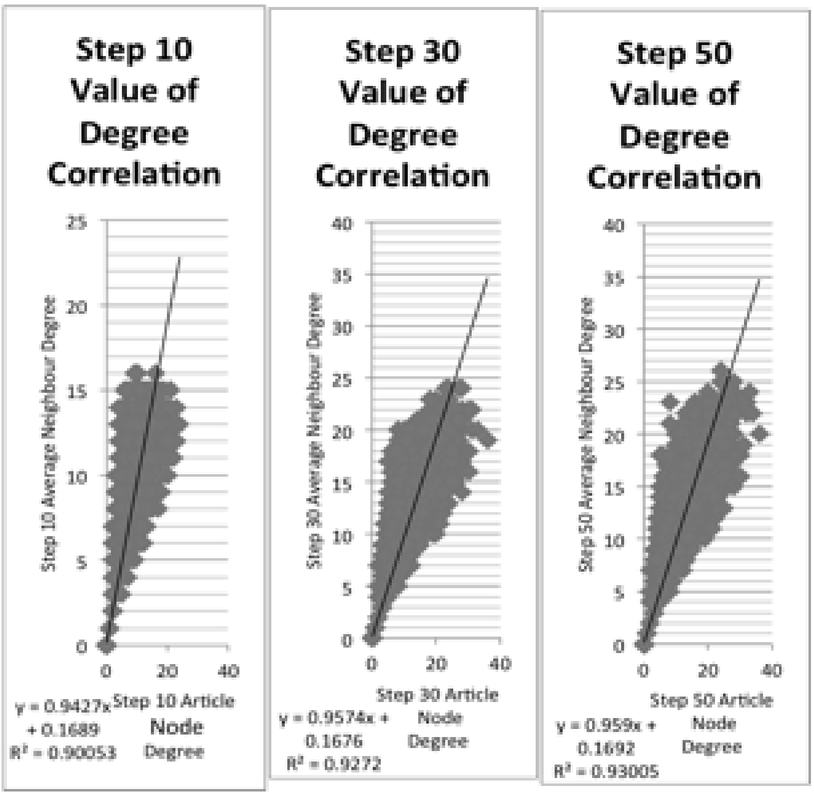




c) Small World coefficient development for user(a) and article(b) networks when GACCT = 8

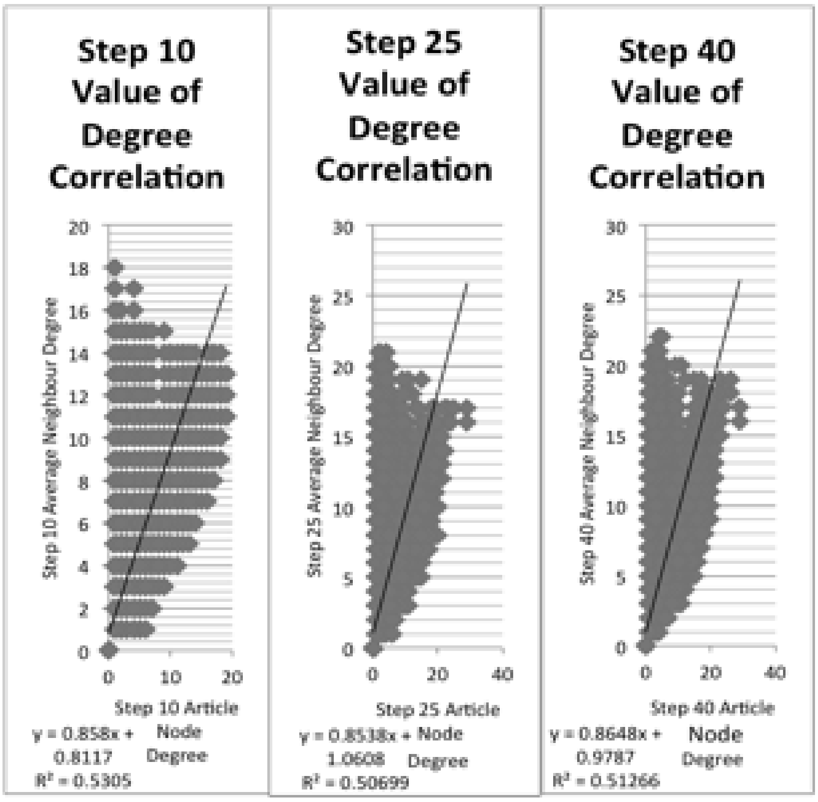
Fig. 28 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GACCT=8.

Article Network Degree Correlation Values



b) Degree distribution development for article network when GACCT = 8

User Network Degree Correlation Values



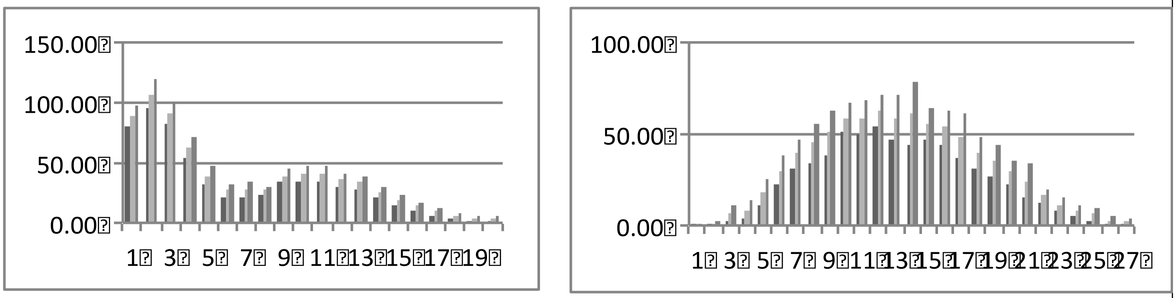
b) Degree distribution development for user network when GACCT = 8

Fig. 29 Degree correlation development for GACCT=8.

Table 9 Article and User Network Degree Distribution for GACCT = 8

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG=STDEV
1	80.40	89.20	98.00
2	94.89	106.75	118.61
3	81.33	90.15	98.97
4	53.41	61.85	70.29
5	31.38	39.00	46.62
6	21.78	27.30	32.82
7	21.54	27.35	33.16
8	23.14	26.95	30.76
9	35.14	39.65	44.16
10	34.61	41.35	48.09
11	33.74	40.20	46.66
12	30.39	35.35	40.31
13	27.04	33.15	39.26
14	20.69	25.70	30.71
15	15.54	19.30	23.06
16	10.10	13.45	16.80
17	6.37	9.25	12.13
18	2.91	5.80	8.69
19	2.17	3.60	5.03
20	0.84	2.85	4.86

Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG=STDEV
1	0.09	0.58	1.08
2	0.03	1.42	2.80
3	2.57	6.83	11.09
4	3.17	8.25	13.33
5	11.53	18.67	25.80
6	21.78	29.75	37.72
7	31.03	39.17	47.30
8	34.33	44.75	55.17
9	38.66	50.92	63.17
10	51.12	58.83	66.55
11	49.35	58.83	68.32
12	54.32	62.58	70.84
13	47.25	59.00	70.75
14	43.82	61.08	78.35
15	46.47	55.50	64.53
16	44.53	53.83	63.13
17	36.17	48.50	60.83
18	31.42	39.92	48.42
19	26.69	35.17	43.64
20	22.42	29.08	35.75
21	15.01	24.50	33.99
22	13.04	16.17	19.30
23	7.51	11.25	14.99
24	4.76	7.83	10.91
25	2.89	6.00	9.11
26	0.32	3.00	5.68
27	0.70	2.50	4.30



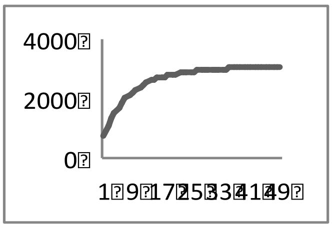
Degree distribution of User(a) and Article(b) networks when GACCT = 8.

Fig. 30 Degree distribution for GACCT=8.

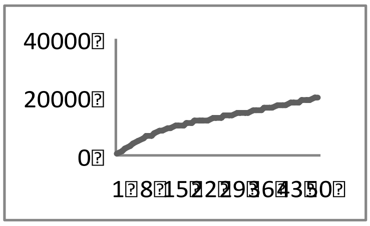
Good Article Connection Count Threshold (GACCT) = 2, n=20 Rep. Results

Table 10 Total Edge Count Development for GACCT = 2

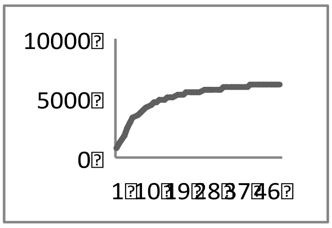
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	691.11	700.00	708.89	287.61	293.40	299.19	665.25	689.80	714.35
Step 2	1030.46	1045.10	1059.74	1008.16	1033.20	1058.24	1260.44	1295.10	1329.76
Step 3	1297.15	1316.80	1336.45	1928.80	1975.30	2021.80	1841.15	1903.40	1965.65
Step 4	1511.43	1534.40	1557.37	2870.47	2939.95	3009.43	2347.14	2433.90	2520.66
Step 5	1692.51	1721.90	1751.29	3807.17	3902.20	3997.23	2784.06	2891.60	2999.14
Step 6	1846.24	1878.80	1911.36	4686.18	4790.50	4894.82	3141.49	3265.00	3388.51
Step 7	1987.72	2018.60	2049.48	5505.53	5606.00	5706.47	3471.87	3597.90	3723.93
Step 8	2102.24	2135.15	2168.06	6240.03	6342.75	6445.47	3733.50	3864.70	3995.90
Step 9	2205.92	2238.80	2271.68	6895.62	7016.40	7137.18	3962.39	4085.70	4209.01
Step 10	2289.99	2326.40	2362.81	7443.58	7605.25	7766.92	4154.31	4278.90	4403.49
Step 11	2369.25	2407.15	2445.05	7982.85	8164.10	8345.35	4325.98	4451.10	4576.22
Step 12	2432.73	2473.00	2513.27	8445.62	8655.20	8864.78	4455.33	4588.00	4720.67
Step 13	2493.48	2534.45	2575.42	8892.61	9116.85	9341.09	4578.48	4715.10	4851.72
Step 14	2545.39	2588.50	2631.61	9313.82	9540.95	9768.08	4692.13	4826.10	4960.07
Step 15	2592.76	2635.75	2678.74	9710.15	9933.65	10157.15	4791.96	4924.10	5056.24
Step 16	2637.98	2679.50	2721.02	10076.13	10295.95	10515.77	4889.49	5018.30	5147.11
Step 17	2673.99	2717.05	2760.11	10402.15	10633.20	10864.25	4972.13	5099.40	5226.67
Step 18	2707.26	2751.05	2794.84	10712.65	10959.85	11207.05	5046.75	5175.00	5303.25
Step 19	2736.12	2780.15	2824.18	11011.25	11266.90	11522.55	5108.90	5241.70	5374.50
Step 20	2760.98	2805.60	2850.22	11288.08	11556.85	11825.62	5173.71	5304.10	5434.49
Step 21	2783.43	2828.65	2873.87	11561.16	11846.35	12131.54	5230.26	5359.20	5488.14
Step 22	2803.88	2850.40	2896.92	11799.12	12110.05	12420.98	5280.00	5409.40	5538.80
Step 23	2822.83	2870.75	2918.67	12054.84	12379.80	12704.76	5330.63	5460.80	5590.97
Step 24	2838.85	2888.60	2938.35	12296.10	12649.85	13003.60	5372.34	5504.00	5635.66
Step 25	2854.47	2904.50	2954.53	12552.72	12901.05	13249.38	5412.87	5546.80	5680.73
Step 26	2867.88	2919.45	2971.02	12810.10	13157.65	13505.20	5453.16	5588.30	5723.44
Step 27	2881.79	2933.00	2984.21	13064.76	13411.05	13757.34	5497.02	5631.60	5766.18
Step 28	2893.26	2944.95	2996.64	13301.78	13651.85	14001.92	5535.11	5669.60	5804.09
Step 29	2904.67	2957.50	3010.33	13539.12	13905.85	14272.58	5573.50	5709.30	5845.10
Step 30	2917.00	2968.50	3020.00	13807.42	14161.85	14516.28	5609.40	5742.90	5876.40
Step 31	2926.48	2977.95	3029.42	14070.53	14423.60	14776.67	5637.59	5776.30	5915.01
Step 32	2935.28	2987.45	3039.62	14329.60	14694.45	15059.30	5670.34	5808.80	5947.26
Step 33	2943.94	2995.60	3047.26	14589.14	14953.70	15318.26	5700.79	5836.80	5972.81
Step 34	2952.26	3003.95	3055.64	14840.79	15221.80	15602.81	5734.97	5868.70	6002.43
Step 35	2959.35	3011.70	3064.05	15107.24	15495.25	15883.26	5761.00	5896.80	6032.60
Step 36	2966.60	3018.60	3070.60	15373.08	15756.65	16140.22	5786.96	5920.50	6054.04
Step 37	2973.59	3025.45	3077.31	15638.52	16033.45	16428.38	5813.12	5948.30	6083.48
Step 38	2979.13	3031.65	3084.17	15906.71	16305.70	16704.69	5835.30	5969.80	6104.30
Step 39	2986.34	3038.60	3090.86	16204.40	16593.25	16982.10	5859.82	5995.20	6130.58
Step 40	2991.80	3044.75	3097.70	16476.98	16872.20	17267.42	5885.15	6020.90	6156.65
Step 41	2997.39	3050.65	3103.91	16745.07	17165.85	17586.63	5911.11	6046.80	6182.49
Step 42	3002.72	3055.85	3108.98	17004.68	17439.80	17874.92	5929.39	6066.30	6203.21
Step 43	3007.76	3061.25	3114.74	17291.35	17740.85	18190.35	5949.15	6087.60	6226.05
Step 44	3013.31	3066.95	3120.59	17582.04	18050.85	18519.66	5970.06	6108.40	6246.74
Step 45	3018.32	3071.80	3125.28	17885.99	18355.65	18825.31	5992.94	6131.30	6269.66
Step 46	3023.10	3076.60	3130.10	18170.77	18669.85	19168.93	6013.20	6150.30	6287.40
Step 47	3027.25	3081.20	3135.15	18452.54	18984.60	19516.66	6032.63	6169.60	6306.57
Step 48	3032.61	3086.20	3139.79	18746.93	19310.95	19874.97	6051.19	6190.00	6328.81
Step 49	3036.55	3090.80	3145.05	19063.29	19645.55	20227.81	6071.53	6210.50	6349.47
Step 50	3040.90	3095.35	3149.80	19386.05	19967.20	20548.35	6090.63	6230.10	6369.57



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

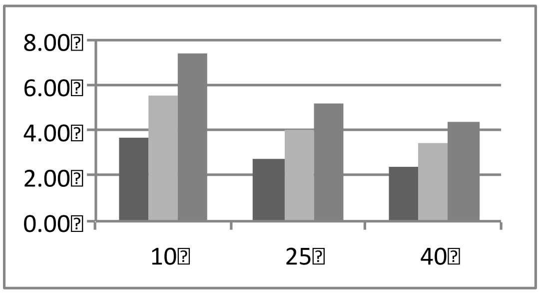
Fig. 31 Network total edge count development for GACCT=2.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.73	5.58	7.44
25	2.81	4.00	5.19
40	2.38	3.40	4.42

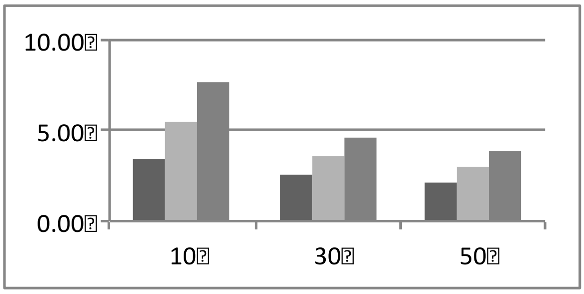
a.



Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.46	5.55	7.65
30	2.53	3.54	4.55
50	2.18	2.99	3.81

b.

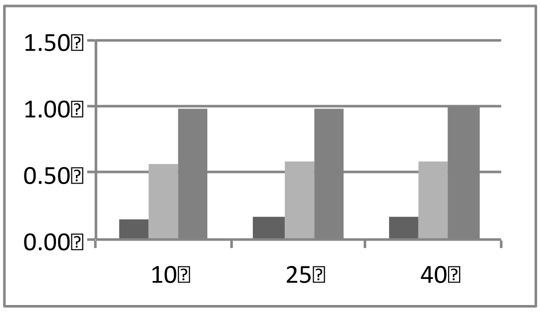


a) Pathlength development for user(a) and article(b) networks when GACCT = 2

Average Clustering Coefficient in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.15	0.56	0.98
25	0.17	0.58	0.99
40	0.18	0.59	1.00

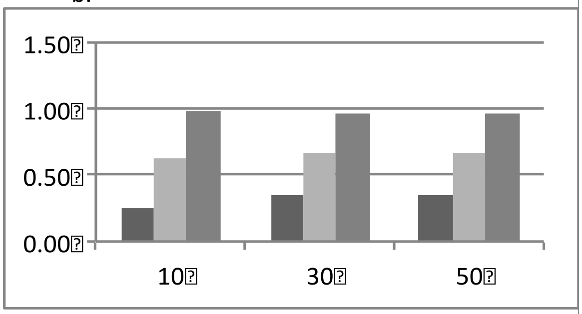
a.



Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.26	0.62	0.98
30	0.35	0.66	0.97
50	0.35	0.66	0.97

b.

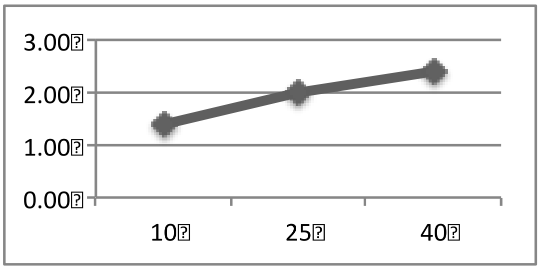


b) Clustering coefficient development for user(a) and article(b) networks when GACCT = 2

Average Small World Coefficient in User Network

Step	AVG
10	1.39
25	2.00
40	2.38

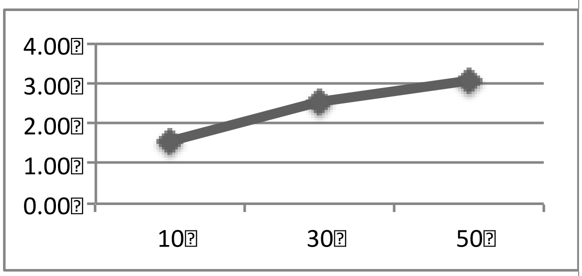
a.



Average Small World Coefficient in Article Network

Step	AVG
10	1.53
30	2.56
50	3.05

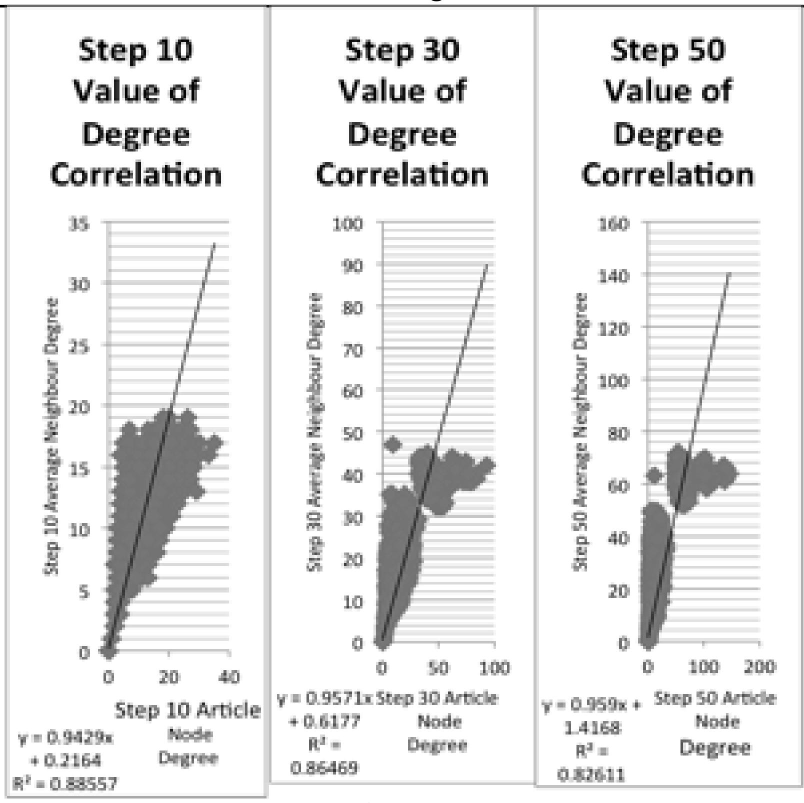
b.



c) Small World coefficient development for user(a) and article(b) networks when GACCT = 2

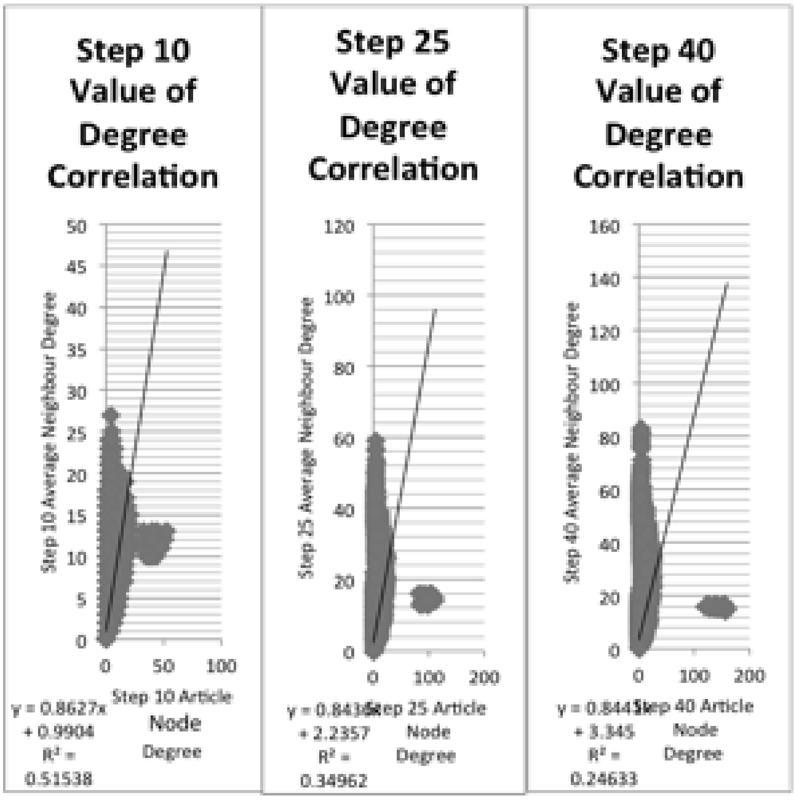
Fig. 32 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GACCT=2.

Article Network Degree Correlation Values



a) Degree distribution development for article network when GACCT = 2

User Network Degree Correlation Values



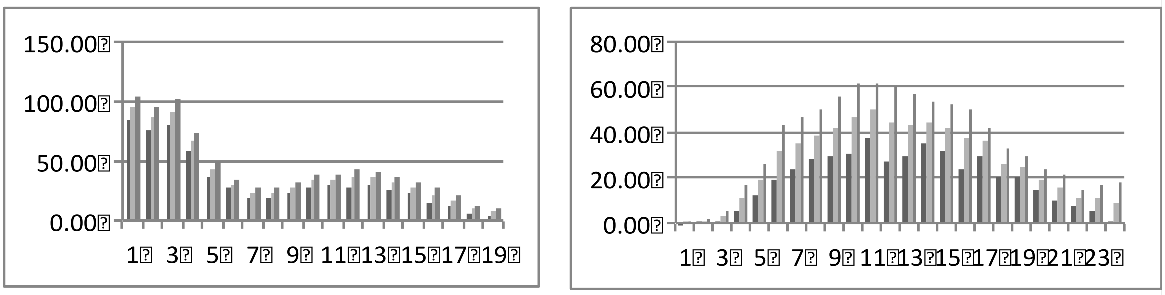
b) Degree distribution development for user network when GACCT = 2

Fig. 33 Degree correlation development for GACCT=2.

Table 11 Article and User Network Degree Distribution for GACCT = 2

Node Degree	User Network User Count Having That Degree: AVG-STDEV	AVG	AVG+STDEV
1	83.68	94.40	105.12
2	76.20	85.55	94.90
3	80.74	91.50	102.26
4	58.33	66.20	74.07
5	35.64	42.15	48.66
6	26.61	30.30	33.99
7	17.82	22.95	28.08
8	18.09	22.55	27.01
9	23.93	27.70	31.47
10	28.34	33.60	38.86
11	28.87	33.80	38.73
12	28.41	35.50	42.59
13	30.32	35.90	41.48
14	25.63	31.25	36.87
15	22.42	26.80	31.18
16	15.16	21.05	26.94
17	12.79	17.10	21.41
18	6.83	9.85	12.87
19	4.29	7.70	11.11

Node Degree	Article Network Article Count Having That Degree: AVG-STDEV	AVG	AVG+STDEV
1	-0.21	0.17	0.54
2	0.29	1.00	1.71
3	0.05	2.92	5.79
4	5.87	11.42	16.96
5	12.02	18.75	25.48
6	19.12	31.25	43.38
7	23.89	35.33	46.78
8	28.21	39.08	49.96
9	28.97	42.33	55.70
10	31.07	46.33	61.60
11	37.59	49.67	61.75
12	27.71	44.08	60.46
13	29.80	43.42	57.03
14	34.67	43.92	53.16
15	32.28	42.25	52.22
16	24.29	37.33	50.38
17	29.86	36.00	42.14
18	20.77	26.58	32.39
19	20.84	25.33	29.83
20	14.57	19.17	23.76
21	9.45	15.58	21.71
22	7.40	11.25	15.10
23	5.36	10.92	16.48
24	0.70	9.08	17.46



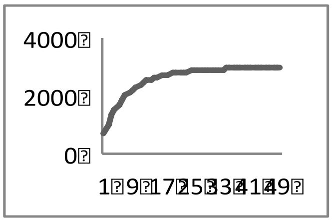
Degree distribution of User(a) and Article(b) networks.

Fig. 34 Degree distribution for GACCT = 2.

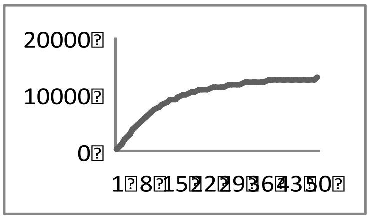
General Interest Percentage (GIP) = 0.005, n=20 Replications Results

Table 12 Total Edge Count Development for GIP = 0.005

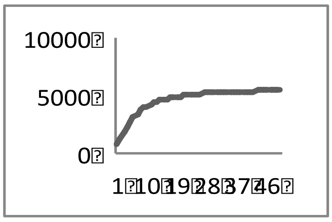
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	689.56	698.20	706.84	291.80	295.45	299.10	654.40	676.90	699.40
Step 2	1027.16	1038.75	1050.34	1019.64	1041.40	1063.16	1189.70	1235.20	1280.70
Step 3	1290.33	1305.60	1320.87	1925.41	1973.80	2022.19	1752.32	1825.35	1898.38
Step 4	1508.41	1526.05	1543.69	2843.92	2930.15	3016.38	2269.69	2351.40	2433.11
Step 5	1690.47	1710.05	1729.63	3728.63	3848.35	3968.07	2700.17	2790.85	2881.53
Step 6	1838.53	1865.60	1892.67	4523.51	4683.10	4842.69	3072.68	3160.90	3249.12
Step 7	1970.41	1998.95	2027.49	5287.29	5455.05	5622.81	3360.80	3451.30	3541.80
Step 8	2090.06	2116.05	2142.04	5969.47	6145.85	6322.23	3613.75	3706.70	3799.65
Step 9	2182.02	2211.75	2241.48	6541.60	6742.20	6942.80	3808.42	3904.60	4000.78
Step 10	2269.28	2301.40	2333.52	7104.58	7299.30	7494.02	3988.27	4090.30	4192.33
Step 11	2341.40	2376.85	2412.30	7580.15	7786.05	7991.95	4129.45	4234.10	4338.75
Step 12	2405.38	2443.30	2481.22	7982.64	8208.05	8433.46	4264.65	4368.40	4472.15
Step 13	2461.61	2501.20	2540.79	8339.32	8585.65	8831.98	4371.56	4477.90	4584.24
Step 14	2509.34	2551.35	2593.36	8673.52	8926.30	9179.08	4461.96	4573.70	4685.44
Step 15	2555.94	2599.70	2643.46	9004.32	9254.30	9504.28	4544.01	4663.50	4782.99
Step 16	2595.44	2640.30	2685.16	9285.34	9549.50	9813.66	4618.38	4738.10	4857.82
Step 17	2628.00	2675.80	2723.60	9528.22	9804.05	10079.88	4678.48	4805.60	4932.72
Step 18	2655.98	2706.20	2756.42	9731.66	10024.00	10316.34	4737.88	4867.90	4997.92
Step 19	2682.21	2732.95	2783.69	9938.40	10221.75	10505.10	4786.78	4918.40	5050.02
Step 20	2704.91	2756.90	2808.89	10115.30	10411.50	10707.70	4828.86	4962.50	5096.14
Step 21	2725.84	2778.30	2830.76	10290.19	10591.20	10892.21	4874.56	5006.20	5137.84
Step 22	2746.41	2800.00	2853.59	10457.82	10763.25	11068.68	4925.75	5054.70	5183.65
Step 23	2764.45	2817.20	2869.95	10600.14	10911.95	11223.76	4956.29	5089.70	5223.11
Step 24	2777.46	2831.85	2886.24	10726.08	11046.90	11367.72	4980.84	5118.40	5255.96
Step 25	2789.70	2845.30	2900.90	10836.81	11171.55	11506.29	5009.28	5145.90	5282.52
Step 26	2803.01	2858.35	2913.69	10969.94	11301.70	11633.46	5038.13	5172.90	5307.67
Step 27	2813.94	2869.45	2924.96	11088.94	11410.35	11731.76	5063.34	5197.60	5331.86
Step 28	2823.88	2879.75	2935.62	11189.74	11525.70	11861.66	5082.66	5218.50	5354.34
Step 29	2832.37	2889.40	2946.43	11282.03	11634.90	11987.77	5096.95	5238.90	5380.85
Step 30	2840.30	2898.30	2956.30	11377.35	11735.05	12092.75	5117.97	5259.80	5401.63
Step 31	2849.10	2906.75	2964.40	11476.94	11838.40	12199.86	5135.14	5277.80	5420.46
Step 32	2856.32	2913.60	2970.88	11562.37	11923.95	12285.53	5148.61	5293.50	5438.39
Step 33	2861.62	2919.85	2978.08	11629.21	12003.45	12377.69	5160.99	5309.80	5458.61
Step 34	2865.69	2924.95	2984.21	11691.86	12076.20	12460.54	5172.70	5323.00	5473.30
Step 35	2869.87	2930.55	2991.23	11756.65	12164.50	12572.35	5182.94	5337.80	5492.66
Step 36	2874.88	2935.65	2996.42	11823.72	12241.15	12658.58	5192.22	5348.20	5504.18
Step 37	2878.83	2939.85	3000.87	11874.60	12306.15	12737.70	5200.01	5356.70	5513.39
Step 38	2882.36	2943.70	3005.04	11917.67	12363.00	12808.33	5207.85	5362.90	5517.95
Step 39	2885.76	2947.45	3009.14	11966.13	12432.40	12898.67	5216.20	5373.80	5531.40
Step 40	2889.39	2950.95	3012.51	12006.80	12483.70	12960.60	5219.02	5380.40	5541.78
Step 41	2891.65	2953.65	3015.65	12037.94	12528.40	13018.86	5223.13	5386.30	5549.47
Step 42	2894.38	2956.60	3018.82	12068.23	12576.80	13085.37	5226.39	5391.00	5555.61
Step 43	2895.75	2958.60	3021.45	12096.24	12613.50	13130.76	5229.74	5395.30	5560.86
Step 44	2897.61	2960.55	3023.49	12115.97	12641.55	13167.13	5232.53	5400.00	5567.47
Step 45	2899.16	2962.10	3025.04	12134.11	12672.30	13210.49	5236.53	5404.70	5572.87
Step 46	2900.73	2963.60	3026.47	12155.28	12701.35	13247.42	5237.74	5408.30	5578.86
Step 47	2902.05	2965.00	3027.95	12173.88	12722.30	13270.72	5238.95	5411.00	5583.05
Step 48	2902.86	2966.25	3029.64	12177.20	12747.05	13316.90	5238.78	5412.90	5587.02
Step 49	2904.49	2968.00	3031.51	12197.53	12768.90	13340.27	5241.54	5416.00	5590.46
Step 50	2905.31	2968.95	3032.59	12203.45	12783.35	13363.25	5242.37	5417.70	5593.03



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

Fig. 35 Network total edge count development for GIP = 0.005.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

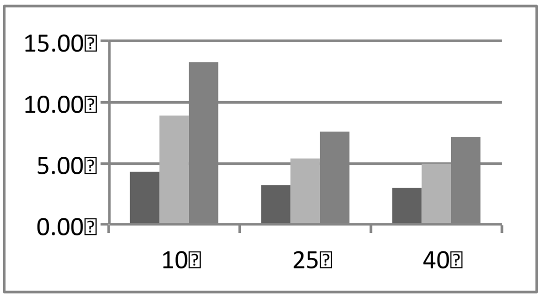
Step	AVG-STDEV	AVG	AVG+STDEV
10	4.31	8.82	13.33
25	3.10	5.38	7.66
40	2.88	5.02	7.16

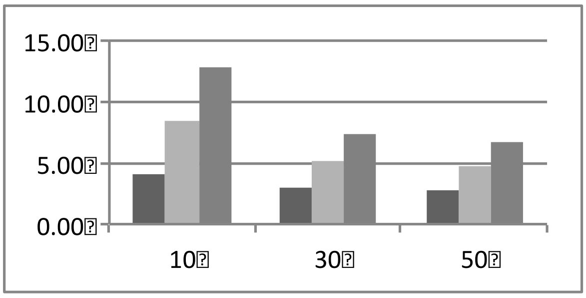
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	4.17	8.50	12.83
30	2.99	5.14	7.29
50	2.78	4.77	6.76

b.





a) Pathlength development for user(a) and article(b) networks when GIP = 0.005.

Average Clustering Coefficient in User Network

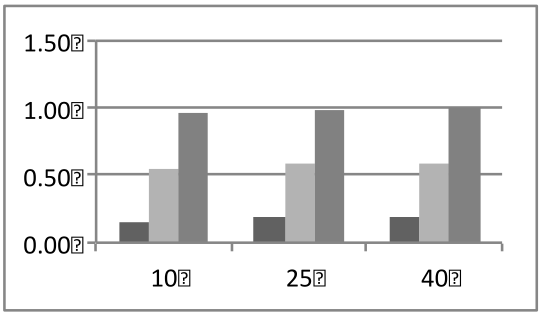
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.14	0.55	0.96
25	0.18	0.59	0.99
40	0.20	0.60	1.00

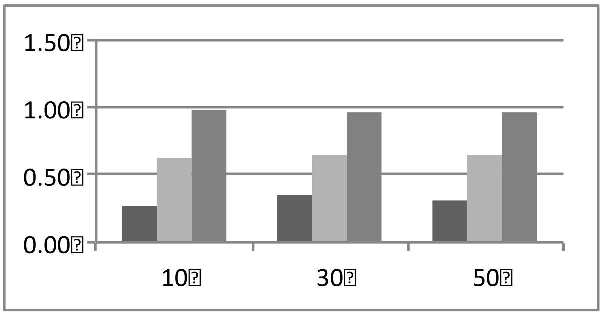
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.27	0.63	0.98
30	0.34	0.66	0.97
50	0.32	0.64	0.96

b.





b) Clustering coefficient development for user(a) and article(b) networks when GIP = 0.005.

Average Small World Coefficient in User Network

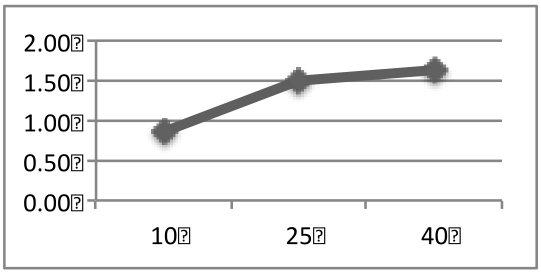
Step	AVG
10	0.86
25	1.50
40	1.64

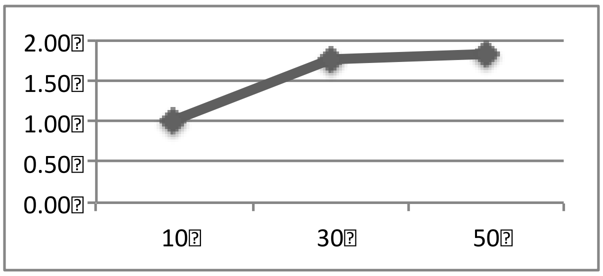
a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.01
30	1.76
50	1.85

b.

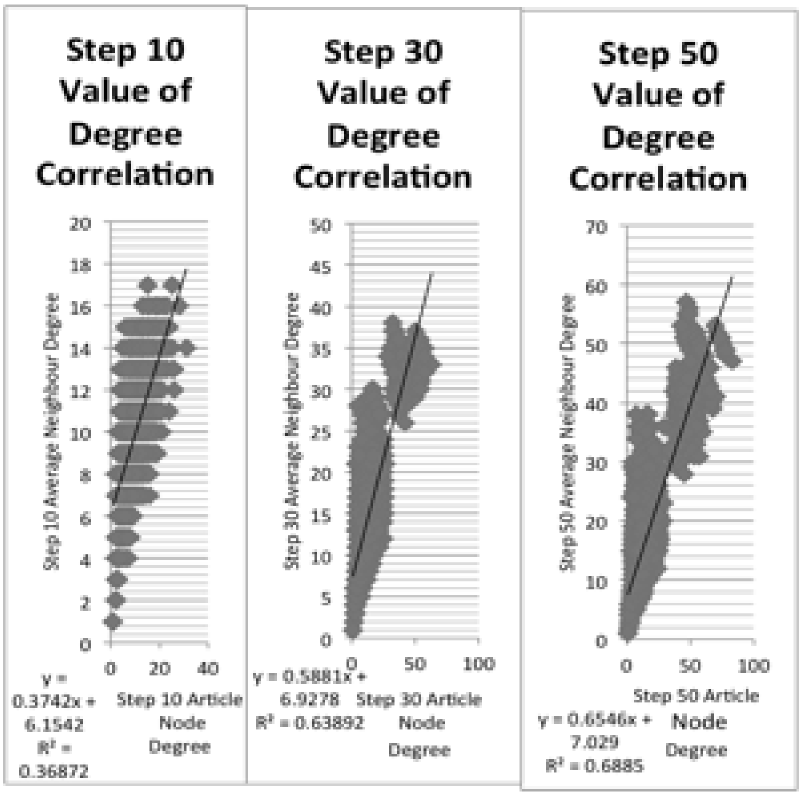




c) Small World coefficient development for user(a) and article(b) networks when GIP = 0.005.

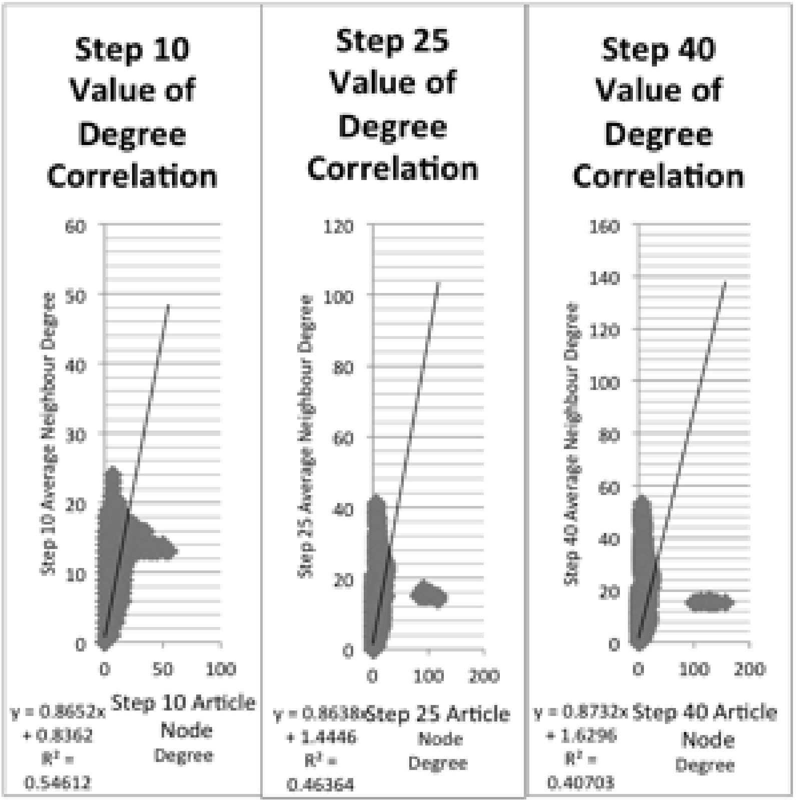
Fig. 36 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GIP = 0.005.

Article Network Degree Correlation Values



a) Degree distribution development for article network when GIP = 0.005.

User Network Degree Correlation Values

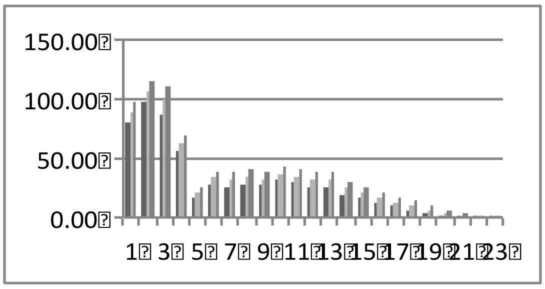


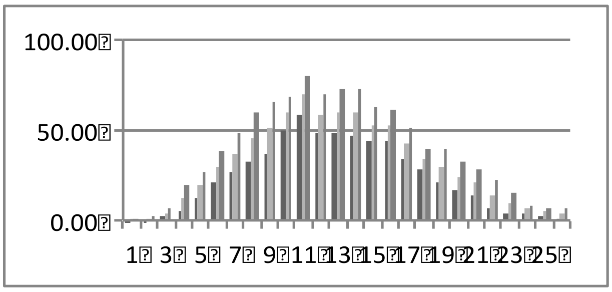
b) Degree distribution development for user network when GIP = 0.005.

Fig . 37 Degree correlation development for GIP = 0.005.

Table 13 Article and User Network Degree Distribution for GIP = 0.005

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG= STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	80.45	89.50	98.55	1	-0.29	0.33	0.96
2	97.93	106.05	114.17	2	-0.04	1.33	2.71
3	87.00	98.70	110.40	3	1.95	4.42	6.88
4	54.96	62.15	69.34	4	5.87	12.67	19.47
5	16.26	20.60	24.94	5	12.01	19.67	27.32
6	28.71	33.45	38.19	6	21.53	29.67	37.80
7	25.59	32.55	39.51	7	26.31	37.00	47.69
8	28.41	34.55	40.69	8	32.44	45.92	59.39
9	27.16	32.80	38.44	9	36.49	51.33	66.18
10	31.32	36.90	42.48	10	50.14	59.25	68.36
11	29.19	34.65	40.11	11	58.74	69.42	80.09
12	25.17	32.05	38.93	12	48.32	59.00	69.68
13	26.29	32.05	37.81	13	48.57	60.42	72.26
14	19.54	24.90	30.26	14	47.32	59.92	72.52
15	17.14	21.00	24.86	15	43.62	52.92	62.21
16	11.35	16.30	21.25	16	43.88	52.25	60.62
17	10.06	13.15	16.24	17	33.77	42.67	51.57
18	6.55	10.25	13.95	18	28.61	34.42	40.23
19	3.66	6.50	9.34	19	20.39	30.25	40.11
20	1.72	3.70	5.68	20	16.24	24.33	32.43
21	0.62	2.15	3.68	21	14.56	21.33	28.11
22	0.01	0.90	1.79	22	6.58	14.50	22.42
23	0.11	0.60	1.09	23	4.10	9.58	15.07
				24	4.43	6.08	7.74
				25	2.19	4.58	6.98
				26	0.72	3.50	6.28





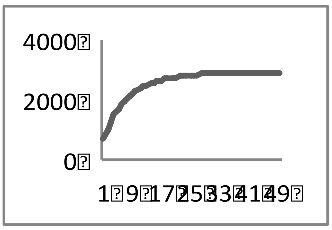
Degree distribution of User(a) and Article(b) networks when GIP = 0.005.

Fig. 38 Degree distribution for GIP = 0.005.

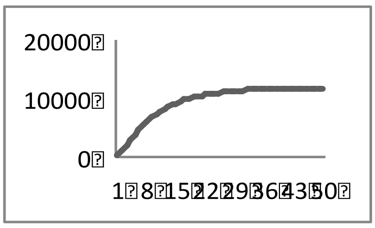
General Interest Percentage (GIP) = 0.015, n=20 Replications Results

Table 14 Total Edge Count Development for GIP = 0.015

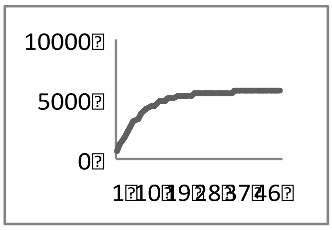
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	668.92	689.55	710.18	275.88	288.85	301.82	616.14	662.35	708.56
Step 2	990.28	1023.10	1055.92	966.78	1014.90	1063.02	1127.46	1208.95	1290.44
Step 3	1238.69	1284.35	1330.01	1822.88	1916.35	2009.82	1649.01	1771.30	1893.59
Step 4	1443.87	1500.25	1556.63	2711.94	2858.40	3004.86	2130.63	2285.95	2441.27
Step 5	1620.77	1684.75	1748.73	3595.99	3777.35	3958.71	2543.75	2743.05	2942.35
Step 6	1767.79	1841.00	1914.21	4413.56	4636.65	4859.74	2894.51	3125.50	3356.49
Step 7	1897.50	1975.55	2053.60	5153.38	5397.80	5642.22	3181.20	3446.65	3712.10
Step 8	2002.55	2091.55	2180.55	5793.01	6092.30	6391.59	3419.89	3727.40	4034.91
Step 9	2090.55	2188.25	2285.95	6334.77	6673.70	7012.63	3623.61	3955.30	4286.99
Step 10	2173.76	2275.05	2376.34	6871.33	7231.35	7591.37	3827.04	4170.10	4513.16
Step 11	2243.48	2350.70	2457.92	7308.96	7706.60	8104.24	3994.37	4347.20	4700.03
Step 12	2304.46	2417.35	2530.24	7699.37	8140.60	8581.83	4124.17	4500.50	4876.83
Step 13	2356.72	2472.70	2588.68	8037.66	8502.50	8967.34	4254.73	4631.70	5008.67
Step 14	2406.40	2525.55	2644.70	8372.25	8855.65	9339.05	4358.53	4754.90	5151.27
Step 15	2448.66	2570.70	2692.74	8677.87	9174.25	9670.63	4450.90	4867.30	5283.70
Step 16	2488.00	2612.10	2736.20	8952.62	9472.40	9992.18	4539.14	4967.05	5394.96
Step 17	2518.86	2646.55	2774.24	9165.70	9711.65	10257.60	4606.62	5049.60	5492.58
Step 18	2546.50	2678.20	2809.90	9363.88	9943.10	10522.32	4669.96	5118.00	5566.04
Step 19	2574.96	2706.90	2838.84	9556.49	10145.20	10733.91	4733.68	5186.50	5639.32
Step 20	2595.62	2731.65	2867.68	9710.89	10333.70	10956.51	4780.09	5243.45	5706.81
Step 21	2615.63	2754.70	2893.77	9853.08	10509.65	11166.22	4820.04	5293.50	5766.96
Step 22	2631.53	2772.60	2913.67	9970.46	10648.70	11326.94	4852.00	5332.35	5812.70
Step 23	2647.77	2789.85	2931.93	10089.87	10786.35	11482.83	4895.77	5376.25	5856.73
Step 24	2660.79	2804.10	2947.41	10190.48	10896.80	11603.12	4929.79	5411.45	5893.11
Step 25	2673.32	2818.50	2963.68	10284.02	11002.75	11721.48	4958.20	5443.40	5928.60
Step 26	2682.35	2829.15	2975.95	10358.66	11088.15	11817.64	4973.63	5463.80	5953.97
Step 27	2693.52	2840.35	2987.18	10441.24	11170.60	11899.96	5000.89	5485.50	5970.11
Step 28	2701.14	2849.75	2998.36	10502.10	11249.75	11997.40	5020.83	5506.75	5992.67
Step 29	2708.73	2857.85	3006.97	10558.16	11318.75	12079.34	5036.69	5525.95	6015.21
Step 30	2717.77	2866.55	3015.33	10624.90	11391.35	12157.80	5059.77	5546.35	6032.93
Step 31	2723.63	2873.15	3022.67	10668.51	11440.45	12212.39	5067.47	5557.35	6047.23
Step 32	2728.35	2879.30	3030.25	10709.24	11486.25	12263.26	5077.56	5571.55	6065.54
Step 33	2732.91	2885.00	3037.09	10747.63	11534.45	12321.27	5087.73	5584.95	6082.17
Step 34	2737.57	2889.90	3042.23	10781.22	11578.60	12375.98	5092.74	5592.55	6092.36
Step 35	2739.84	2893.45	3047.06	10802.54	11611.65	12420.76	5098.72	5600.75	6102.78
Step 36	2744.59	2898.70	3052.81	10837.43	11656.50	12475.57	5109.78	5613.00	6116.22
Step 37	2747.16	2901.90	3056.64	10857.30	11682.95	12508.60	5114.30	5620.10	6125.90
Step 38	2750.24	2904.55	3058.86	10884.55	11707.25	12529.95	5118.36	5624.40	6130.44
Step 39	2752.07	2907.20	3062.33	10898.54	11729.40	12560.26	5122.11	5629.90	6137.69
Step 40	2754.32	2909.30	3064.28	10918.89	11750.45	12582.01	5129.51	5635.95	6142.39
Step 41	2756.39	2911.65	3066.91	10935.08	11769.35	12603.62	5135.06	5641.75	6148.44
Step 42	2759.20	2914.20	3069.20	10950.80	11790.75	12630.70	5140.48	5646.30	6152.12
Step 43	2760.63	2915.50	3070.37	10965.20	11802.35	12639.50	5144.99	5649.10	6153.21
Step 44	2761.36	2916.55	3071.74	10972.35	11812.55	12652.75	5146.68	5651.60	6156.52
Step 45	2763.06	2917.90	3072.74	10982.79	11825.40	12668.01	5151.54	5655.35	6159.16
Step 46	2764.39	2919.20	3074.01	10994.51	11835.30	12676.09	5154.25	5657.75	6161.25
Step 47	2765.75	2920.65	3075.55	11005.82	11849.25	12692.68	5159.61	5661.10	6162.59
Step 48	2766.74	2921.85	3076.96	11012.89	11859.65	12706.41	5163.94	5664.40	6164.86
Step 49	2767.58	2922.70	3077.82	11019.43	11866.45	12713.47	5164.19	5664.90	6165.61
Step 50	2768.05	2923.50	3078.95	11021.68	11871.35	12721.02	5164.78	5665.80	6166.82



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

Fig. 39 Network total edge count development for GIP = 0.015.

Key Network Metrics for Article and User Networks

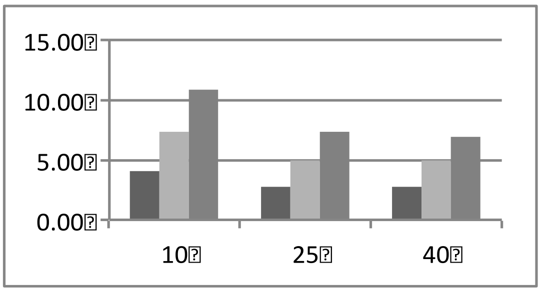
Average Pathlength in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	4.09	7.42	10.75
25	2.78	5.02	7.27
40	2.83	4.87	6.91

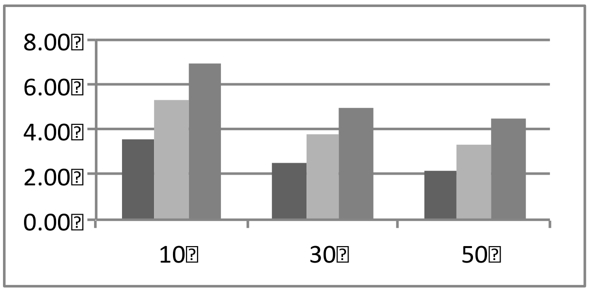
Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.56	5.29	7.01
30	2.52	3.77	5.03
50	2.19	3.32	4.46

a.



b.



a) Pathlength development for user(a) and article(b) networks when GIP = 0.015

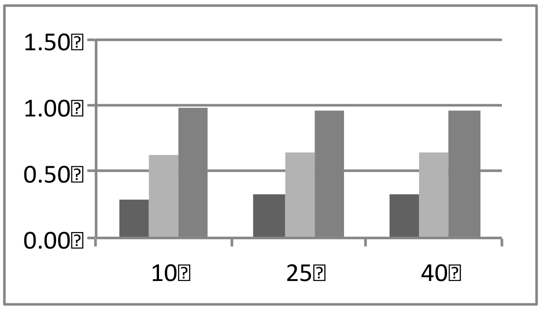
Average Clustering Coefficient in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.28	0.64	0.99
25	0.32	0.65	0.97
40	0.32	0.64	0.96

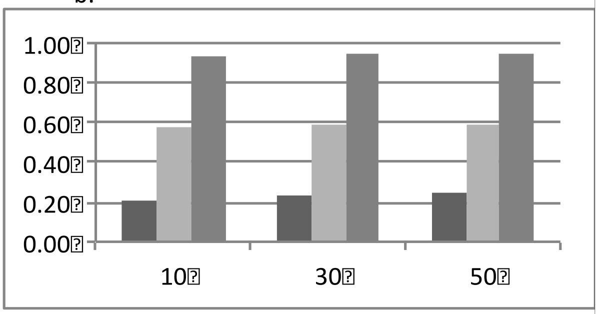
Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.21	0.57	0.93
30	0.24	0.59	0.94
50	0.24	0.59	0.95

a.



b.



b) Clustering coefficient development for user(a) and article(b) networks when GIP = 0.015

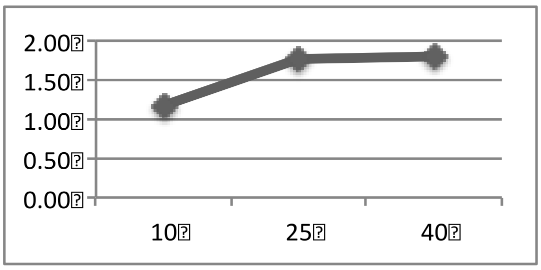
Average Small World Coefficient in User Network

Step	AVG
10	1.18
25	1.77
40	1.81

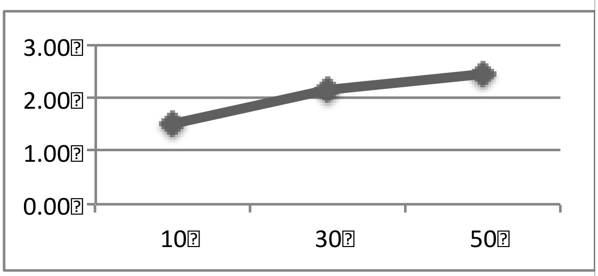
Average Small World Coefficient in Article Network

Step	AVG
10	1.49
30	2.15
50	2.46

a.



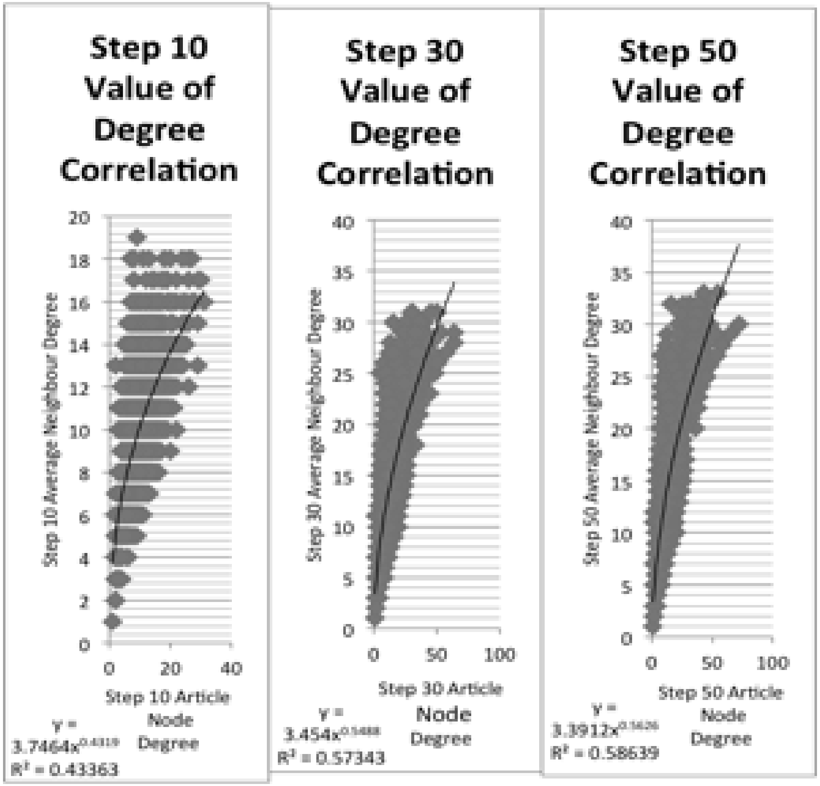
b.



c) Small World coefficient development for user(a) and article(b) networks when GIP = 0.015

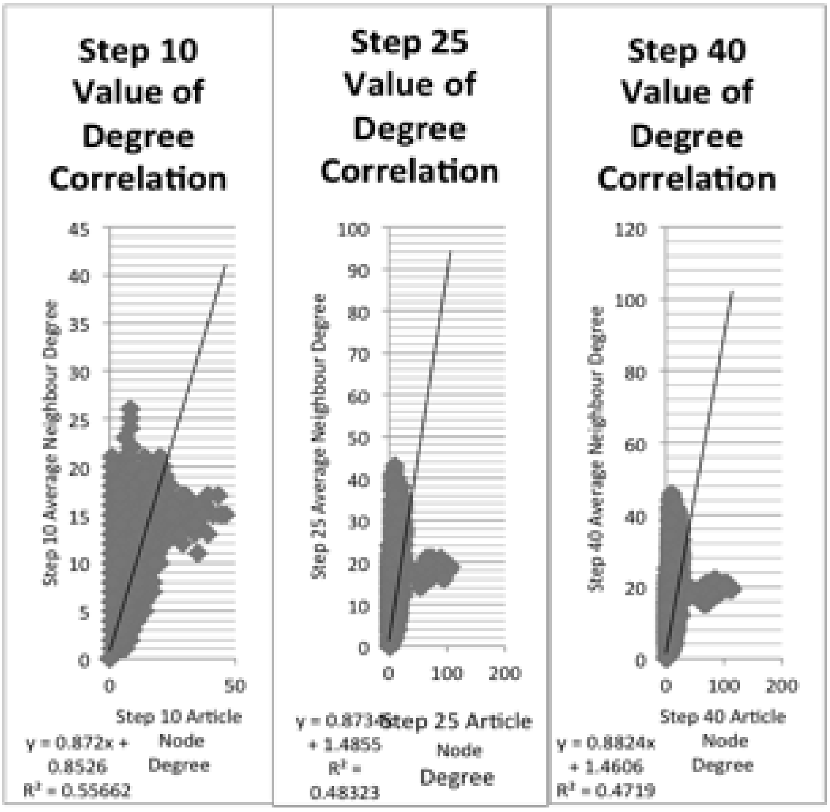
Fig. 40 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GIP = 0.015.

Article Network Degree Correlation Values



a) Degree distribution development for article network when GIP = 0.015

User Network Degree Correlation Values

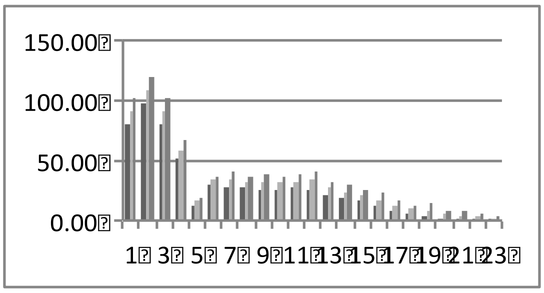


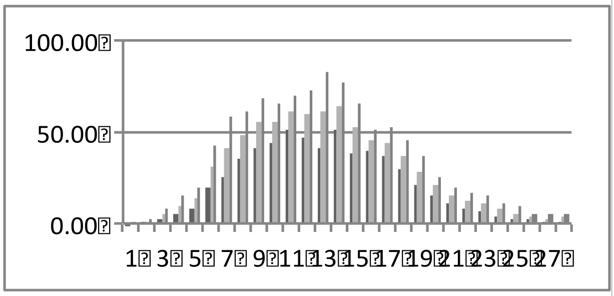
b) Degree distribution development for user network when GIP = 0.015

Fig. 41 Degree correlation development for GIP = 0.015.

Table 15 Article and User Network Degree Distribution for GIP = 0.015

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG+STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG+STDEV
1	79.33	90.25	101.17	1	-0.29	0.33	0.96
2	96.95	108.40	119.85	2	0.24	1.25	2.26
3	81.10	91.20	101.30	3	2.24	5.08	7.93
4	51.41	59.10	66.79	4	5.28	10.25	15.22
5	12.69	15.75	18.81	5	7.96	13.75	19.54
6	30.20	33.60	37.00	6	20.14	31.33	42.53
7	27.44	33.95	40.46	7	24.72	41.42	58.12
8	27.65	32.15	36.65	8	35.67	48.33	61.00
9	26.36	32.50	38.64	9	41.11	54.92	68.73
10	26.46	31.95	37.44	10	44.26	55.25	66.24
11	27.03	32.50	37.97	11	51.90	61.08	70.27
12	26.07	33.40	40.73	12	46.88	59.67	72.46
13	22.06	27.55	33.04	13	40.61	61.50	82.39
14	18.00	23.95	29.90	14	50.71	64.17	77.63
15	17.57	21.75	25.93	15	38.86	52.17	65.47
16	12.90	17.55	22.20	16	39.66	45.08	50.51
17	8.68	13.20	17.72	17	36.25	44.08	51.92
18	6.41	9.75	13.09	18	29.49	37.17	44.85
19	4.25	8.95	13.65	19	21.06	28.75	36.44
20	2.40	5.60	8.80	20	15.99	20.75	25.51
21	1.30	4.25	7.20	21	11.42	15.25	19.08
22	0.61	2.85	5.09	22	8.81	12.50	16.19
23	0.16	2.00	3.84	23	6.27	10.58	14.89
				24	4.46	7.58	10.70
				25	1.85	5.75	9.65
				26	1.76	3.42	5.07
				27	0.87	3.08	5.30
				28	0.66	3.25	5.84





Degree distribution of User(a) and Article(b) networks when GIP = 0.015.

Fig. 42 Degree distribution for GIP = 0.015.

Good Article Multiplier (GAM) = 7.5, n=20 Replications Results

Table 16 Total Edge Count Development for GAM =7.5

	Affiliation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	685.68	694.10	702.52	283.89	291.40	298.91	636.24	672.00	707.76
Step 2	1019.79	1034.95	1050.11	1006.28	1030.05	1053.82	1186.96	1252.55	1318.14
Step 3	1276.43	1296.70	1316.97	1903.21	1949.10	1994.99	1749.00	1838.95	1928.90
Step 4	1493.12	1516.35	1539.58	2831.25	2911.25	2991.25	2264.57	2371.25	2477.93
Step 5	1676.53	1700.10	1723.67	3736.10	3829.30	3922.50	2706.75	2816.05	2925.35
Step 6	1829.20	1855.70	1882.20	4562.48	4676.00	4789.52	3064.67	3184.35	3304.03
Step 7	1959.62	1989.25	2018.88	5303.89	5442.40	5580.91	3344.47	3475.60	3606.73
Step 8	2072.93	2104.50	2136.07	5941.90	6112.10	6282.30	3561.33	3708.10	3854.87
Step 9	2171.68	2204.40	2237.12	6557.94	6733.90	6909.86	3748.74	3899.50	4050.26
Step 10	2250.38	2286.95	2323.52	7032.61	7237.50	7442.39	3897.96	4051.50	4205.04
Step 11	2319.01	2359.70	2400.39	7461.19	7690.15	7919.11	4021.67	4184.20	4346.73
Step 12	2382.17	2423.95	2465.73	7838.90	8086.25	8333.60	4128.65	4292.30	4455.95
Step 13	2434.48	2478.80	2523.12	8158.31	8410.20	8662.09	4215.22	4383.90	4552.58
Step 14	2480.01	2526.95	2573.89	8429.01	8713.45	8997.89	4285.20	4460.80	4636.40
Step 15	2519.97	2566.50	2613.03	8675.77	8968.05	9260.33	4353.49	4527.90	4702.31
Step 16	2557.48	2604.70	2651.92	8929.65	9217.95	9506.25	4414.27	4587.10	4759.93
Step 17	2590.46	2638.60	2686.74	9132.72	9436.60	9740.48	4464.71	4640.80	4816.89
Step 18	2618.48	2667.65	2716.82	9302.56	9622.90	9943.24	4510.19	4686.30	4862.41
Step 19	2641.82	2692.45	2743.08	9457.63	9780.60	10103.57	4548.84	4724.40	4899.96
Step 20	2663.83	2714.55	2765.27	9598.61	9919.00	10239.39	4583.12	4757.80	4932.48
Step 21	2682.44	2734.55	2786.66	9708.68	10040.20	10371.72	4615.02	4788.70	4962.38
Step 22	2700.84	2753.35	2805.86	9831.10	10165.80	10500.50	4639.78	4814.70	4989.62
Step 23	2716.45	2769.10	2821.75	9928.13	10270.90	10613.67	4660.46	4837.70	5014.94
Step 24	2730.82	2785.10	2839.38	10032.19	10381.20	10730.21	4682.83	4863.00	5043.17
Step 25	2742.71	2796.70	2850.69	10098.63	10452.55	10806.47	4701.61	4881.30	5060.99
Step 26	2753.66	2808.35	2863.04	10165.76	10531.95	10898.14	4715.10	4896.60	5078.10
Step 27	2760.76	2816.80	2872.84	10211.51	10587.15	10962.79	4728.88	4910.30	5091.72
Step 28	2769.02	2826.10	2883.18	10269.63	10647.85	11026.07	4740.53	4923.20	5105.87
Step 29	2777.68	2835.30	2892.92	10325.20	10708.70	11092.20	4750.42	4933.80	5117.18
Step 30	2784.22	2841.95	2899.68	10372.92	10757.45	11141.98	4759.33	4942.00	5124.67
Step 31	2789.93	2847.55	2905.17	10412.60	10800.45	11188.30	4767.41	4950.10	5132.79
Step 32	2794.15	2852.55	2910.95	10434.88	10830.65	11226.42	4773.62	4957.40	5141.18
Step 33	2800.02	2859.00	2917.98	10471.38	10872.90	11274.42	4782.44	4966.60	5150.76
Step 34	2804.54	2863.70	2922.86	10496.14	10903.50	11310.86	4788.82	4972.40	5155.98
Step 35	2808.39	2867.85	2927.31	10516.85	10929.85	11342.85	4793.91	4977.90	5161.89
Step 36	2812.00	2871.50	2931.00	10546.89	10958.95	11371.01	4800.77	4983.50	5166.23
Step 37	2815.46	2874.50	2933.54	10567.43	10979.10	11390.77	4806.35	4988.00	5169.65
Step 38	2818.47	2877.45	2936.43	10582.03	10998.60	11415.17	4810.20	4991.20	5172.20
Step 39	2821.39	2880.35	2939.31	10610.13	11022.35	11434.57	4814.55	4994.30	5174.05
Step 40	2824.08	2883.00	2941.92	10627.05	11041.35	11455.65	4819.54	4998.40	5177.26
Step 41	2826.03	2884.90	2943.77	10639.16	11054.70	11470.24	4823.05	5001.10	5179.15
Step 42	2827.01	2886.60	2946.19	10644.63	11064.20	11483.77	4825.33	5003.00	5180.67
Step 43	2828.43	2888.30	2948.17	10656.75	11075.65	11494.55	4827.09	5004.70	5182.31
Step 44	2830.15	2889.90	2949.65	10671.44	11088.85	11506.26	4828.43	5006.70	5184.97
Step 45	2831.38	2891.20	2951.02	10683.80	11100.75	11517.70	4830.03	5008.20	5186.37
Step 46	2832.57	2892.65	2952.73	10692.48	11109.80	11527.12	4830.86	5009.90	5188.94
Step 47	2833.41	2893.70	2953.99	10695.97	11117.05	11538.13	4831.87	5010.90	5189.93
Step 48	2834.54	2894.70	2954.86	10707.50	11127.15	11546.80	4833.26	5012.20	5191.14
Step 49	2835.13	2895.15	2955.17	10710.13	11130.55	11550.97	4833.68	5012.60	5191.52
Step 50	2835.77	2895.95	2956.13	10716.96	11136.95	11556.94	4835.03	5013.70	5192.37

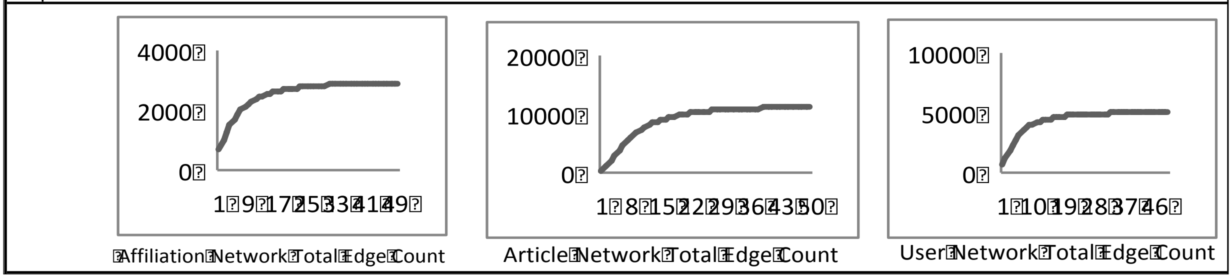


Fig. 43 Network total edge count development for GAM = 7.5.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

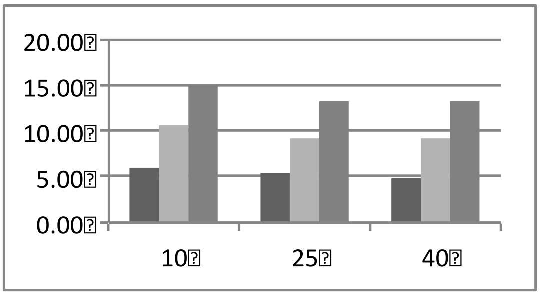
Step	AVG-STDEV	AVG	AVG+STDEV
10	6.15	10.66	15.16
25	5.31	9.36	13.41
40	4.99	9.18	13.37

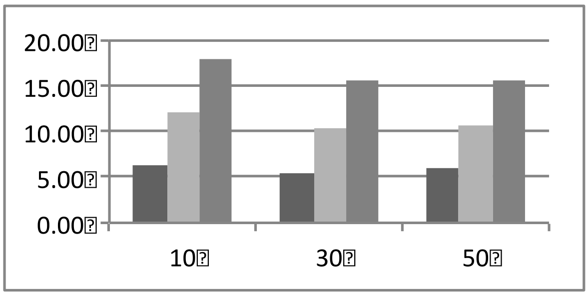
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	6.39	12.21	18.03
30	5.49	10.51	15.53
50	5.89	10.71	15.53

b.





a) Pathlength development for user(a) and article(b) networks when GAM = 7.5

Average Clustering Coefficient in User Network

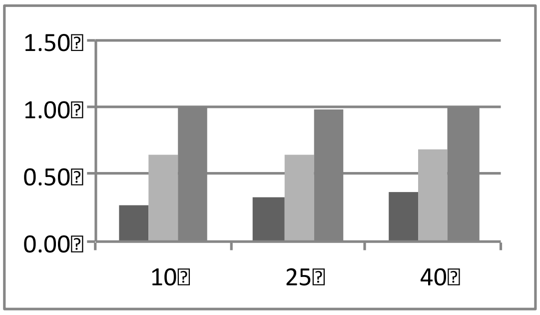
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.27	0.64	1.00
25	0.32	0.65	0.98
40	0.38	0.69	1.00

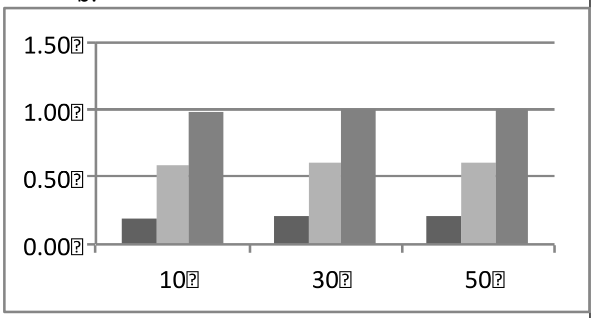
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.19	0.59	0.99
30	0.22	0.61	1.01
50	0.22	0.61	1.00

b.





b) Clustering coefficient development for user(a) and article(b) networks when GAM = 7.5

Average Small World Coefficient in User Network

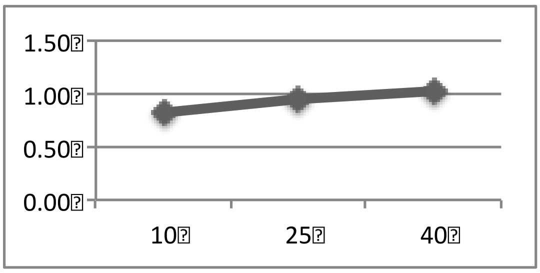
Step	AVG
10	0.82
25	0.96
40	1.03

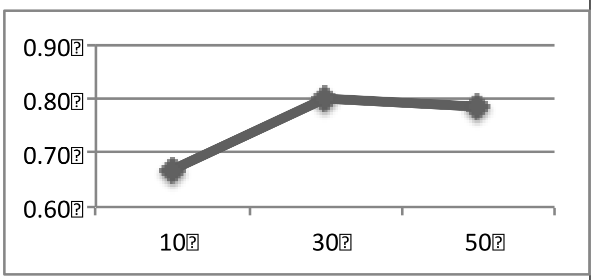
a.

Average Small World Coefficient in Article Network

Step	AVG
10	0.67
30	0.80
50	0.78

b.

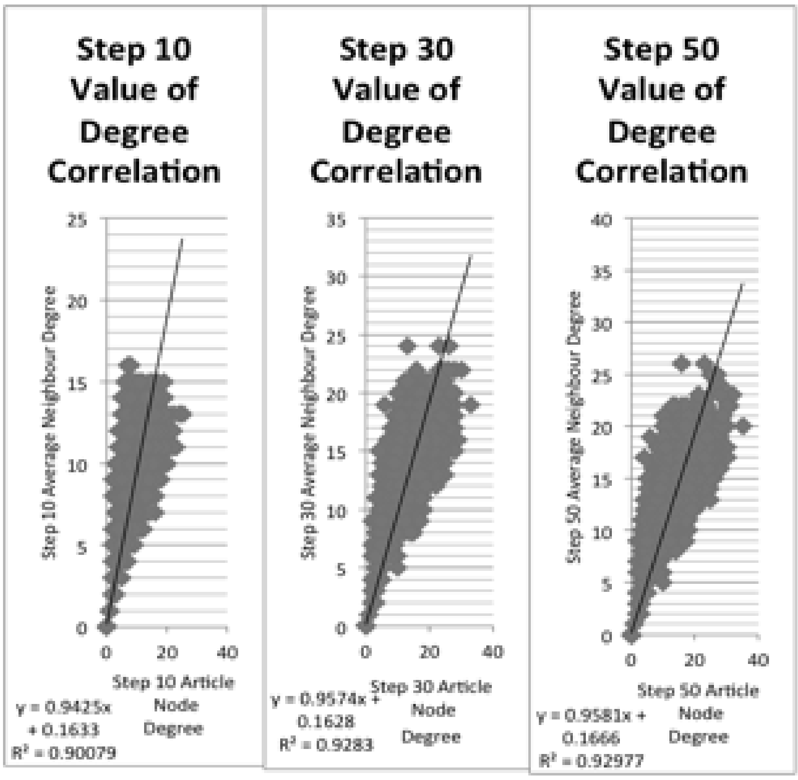




c) Small World coefficient development for user(a) and article(b) networks when GAM = 7.5

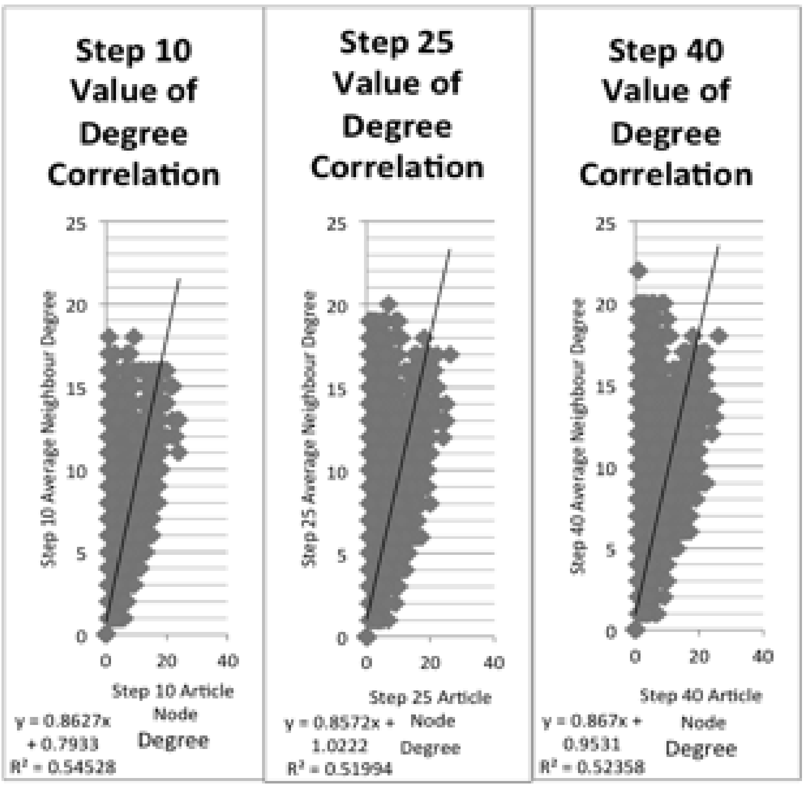
Fig. 44 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GAM=7.5.

Article Network Degree Correlation Values



a) Degree distribution development for article network when GAM = 7.5

User Network Degree Correlation Values

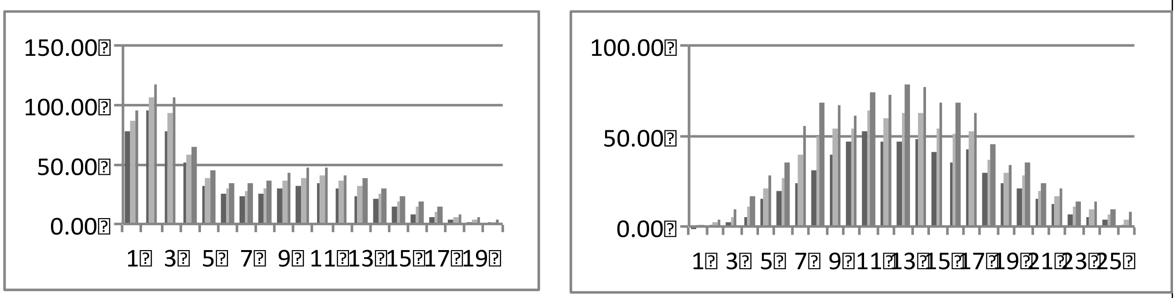


b) Degree distribution development for user network when GAM = 7.5

Fig .45 Degree correlation development for GAM = 7.5.

Table 17 Article and User Network Degree Distribution for GAM = 7.5

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG=STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG=STDEV
1	77.96	86.50	95.04	1	-0.22	0.42	1.06
2	95.02	106.55	118.08	2	0.42	2.25	4.08
3	78.52	92.85	107.18	3	2.32	5.67	9.02
4	52.59	59.10	65.61	4	5.93	11.25	16.57
5	31.45	38.45	45.45	5	15.02	21.75	28.48
6	25.04	30.15	35.26	6	19.79	27.33	34.88
7	23.17	28.65	34.13	7	24.34	40.00	55.66
8	24.82	30.45	36.08	8	31.35	49.58	67.82
9	30.62	36.45	42.28	9	39.92	53.42	66.91
10	31.90	39.25	46.60	10	46.74	53.75	60.76
11	34.92	40.85	46.78	11	52.27	63.42	74.56
12	30.32	35.85	41.38	12	47.07	59.83	72.60
13	24.32	31.40	38.48	13	46.54	62.17	77.79
14	20.43	25.40	30.37	14	48.92	63.08	77.25
15	14.29	18.80	23.31	15	40.73	54.50	68.27
16	9.00	13.80	18.60	16	34.90	51.58	68.26
17	6.60	10.15	13.70	17	43.18	53.17	63.16
18	3.27	6.05	8.83	18	29.11	37.25	45.39
19	1.59	4.00	6.41	19	24.38	29.50	34.62
20	0.38	2.25	4.12	20	20.55	27.75	34.95
				21	14.89	19.25	23.61
				22	11.95	16.17	20.39
				23	6.93	10.33	13.73
				24	5.31	9.92	14.53
				25	3.50	6.83	10.17
				26	1.15	4.50	7.85



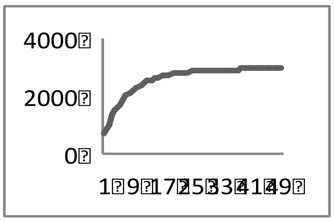
Degree distribution of User(a) and Article(b) networks when GAM = 7.5.

Fig. 46 Degree distribution for GAM = 7.5.

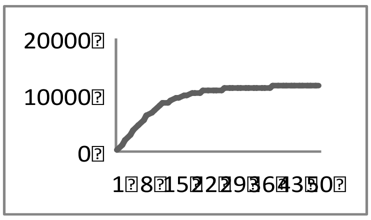
Good Article Multiplier (GAM) = 5, n=20 Replications Results

Table 18 Total Edge Count Development for GAM = 5

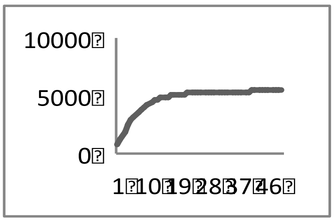
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	688.51	698.10	707.69	288.14	295.35	302.56	646.92	678.80	710.68
Step 2	1024.52	1041.80	1059.08	1012.50	1047.25	1082.00	1208.29	1266.45	1324.61
Step 3	1286.33	1310.35	1334.37	1920.05	1985.60	2051.15	1788.46	1865.10	1941.74
Step 4	1503.71	1531.65	1559.59	2862.95	2949.50	3036.05	2314.69	2400.30	2485.91
Step 5	1687.72	1720.35	1752.98	3776.19	3898.85	4021.51	2766.95	2864.10	2961.25
Step 6	1837.71	1878.10	1918.49	4604.57	4766.40	4928.23	3124.49	3243.60	3362.71
Step 7	1973.47	2014.35	2055.23	5389.58	5552.25	5714.92	3446.77	3573.30	3699.83
Step 8	2084.85	2130.70	2176.55	6077.23	6251.65	6426.07	3717.10	3850.30	3983.50
Step 9	2184.42	2232.35	2280.28	6699.16	6879.45	7059.74	3933.02	4080.20	4227.38
Step 10	2269.67	2322.00	2374.33	7246.62	7445.70	7644.78	4129.76	4289.30	4448.84
Step 11	2346.13	2400.65	2455.17	7748.97	7956.25	8163.53	4299.68	4463.80	4627.92
Step 12	2409.93	2467.60	2525.27	8192.37	8405.25	8618.13	4438.59	4607.80	4777.01
Step 13	2463.34	2524.55	2585.76	8558.78	8782.35	9005.92	4538.86	4715.10	4891.34
Step 14	2511.47	2572.30	2633.13	8877.11	9089.95	9302.79	4633.19	4804.70	4976.21
Step 15	2551.44	2614.35	2677.26	9135.43	9368.10	9600.77	4699.96	4880.00	5060.04
Step 16	2591.31	2655.70	2720.09	9406.46	9641.50	9876.54	4771.68	4949.10	5126.52
Step 17	2626.93	2691.60	2756.27	9618.42	9867.70	10116.98	4834.14	5011.70	5189.26
Step 18	2655.25	2719.50	2783.75	9787.54	10047.20	10306.86	4884.27	5060.50	5236.73
Step 19	2680.29	2743.95	2807.61	9961.89	10222.90	10483.91	4926.10	5101.20	5276.30
Step 20	2703.54	2766.95	2830.36	10107.80	10369.30	10630.80	4966.50	5140.00	5313.50
Step 21	2722.49	2785.70	2848.91	10238.95	10496.50	10754.05	5003.30	5171.50	5339.70
Step 22	2739.33	2803.15	2866.97	10350.93	10614.10	10877.27	5027.33	5197.80	5368.27
Step 23	2753.03	2818.30	2883.57	10436.02	10713.10	10990.18	5054.11	5224.10	5394.09
Step 24	2767.17	2831.95	2896.73	10531.47	10810.55	11089.63	5079.45	5245.50	5411.55
Step 25	2779.60	2845.10	2910.60	10609.93	10893.55	11177.17	5099.15	5264.20	5429.25
Step 26	2792.23	2857.05	2921.87	10689.89	10977.00	11264.11	5119.17	5283.10	5447.03
Step 27	2801.16	2866.75	2932.34	10753.52	11039.40	11325.28	5135.33	5298.10	5460.87
Step 28	2811.29	2877.45	2943.61	10818.69	11106.75	11394.81	5149.72	5313.50	5477.28
Step 29	2818.34	2884.85	2951.36	10861.41	11151.80	11442.19	5158.94	5323.90	5488.86
Step 30	2825.13	2891.85	2958.57	10912.33	11198.80	11485.27	5167.24	5332.80	5498.36
Step 31	2831.67	2898.50	2965.33	10957.41	11243.35	11529.29	5176.11	5342.70	5509.29
Step 32	2836.83	2904.65	2972.47	10986.04	11284.85	11583.66	5183.21	5350.70	5518.19
Step 33	2841.36	2909.20	2977.04	11017.26	11315.20	11613.14	5191.10	5356.90	5522.70
Step 34	2845.75	2914.05	2982.35	11043.43	11345.90	11648.37	5197.68	5363.70	5529.72
Step 35	2849.33	2918.65	2987.97	11069.71	11378.60	11687.49	5205.74	5371.50	5537.26
Step 36	2852.34	2922.00	2991.66	11096.34	11405.40	11714.46	5208.91	5375.30	5541.69
Step 37	2855.73	2925.20	2994.67	11115.10	11425.15	11735.20	5212.17	5378.80	5545.43
Step 38	2857.69	2927.55	2997.41	11124.86	11441.85	11758.84	5215.03	5381.90	5548.77
Step 39	2860.88	2930.60	3000.32	11145.29	11462.20	11779.11	5217.98	5385.80	5553.62
Step 40	2862.98	2932.85	3002.72	11166.91	11482.50	11798.09	5220.91	5389.00	5557.09
Step 41	2864.91	2934.90	3004.89	11186.83	11498.90	11810.97	5224.38	5392.10	5559.82
Step 42	2867.31	2937.05	3006.79	11201.80	11513.05	11824.30	5228.27	5395.40	5562.53
Step 43	2868.88	2938.55	3008.22	11218.80	11525.75	11832.70	5230.35	5398.20	5566.05
Step 44	2869.67	2939.95	3010.23	11217.90	11535.00	11852.10	5232.02	5400.10	5568.18
Step 45	2870.48	2941.20	3011.92	11226.92	11543.60	11860.28	5232.81	5401.40	5569.99
Step 46	2872.35	2942.45	3012.55	11235.92	11552.35	11868.78	5234.70	5402.50	5570.30
Step 47	2873.40	2943.45	3013.50	11238.97	11560.30	11881.63	5235.81	5403.70	5571.59
Step 48	2874.88	2944.55	3014.22	11251.47	11571.60	11891.73	5237.69	5405.10	5572.51
Step 49	2876.11	2945.85	3015.59	11261.85	11581.55	11901.25	5238.97	5406.80	5574.63
Step 50	2876.58	2946.70	3016.82	11264.27	11587.15	11910.03	5239.53	5408.00	5576.47



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

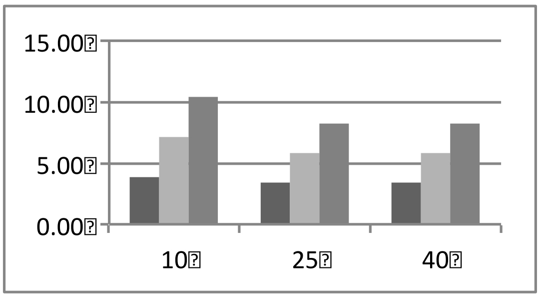
Fig. 47 Network total edge count development for GAM = 5.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.93	7.14	10.35
25	3.49	5.86	8.24
40	3.47	5.84	8.21

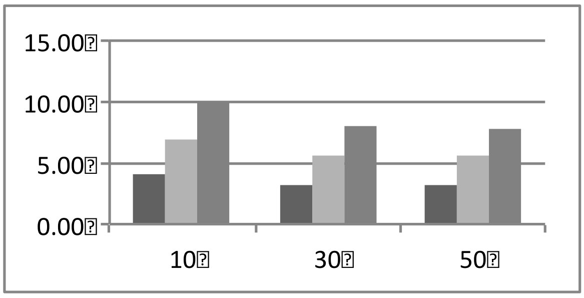
a.



Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	4.06	7.02	9.98
30	3.25	5.61	7.97
50	3.26	5.56	7.87

b.

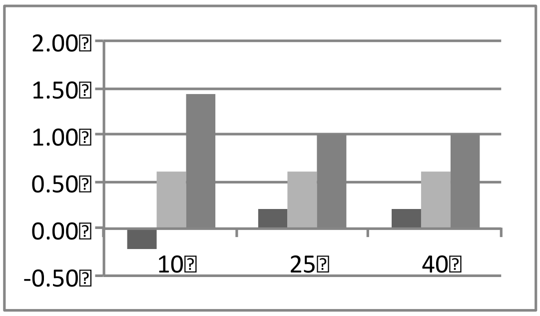


a) Pathlength development for user(a) and article(b) networks when GAM = 5.

Average Clustering Coefficient in User Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	-0.22	0.61	1.44
25	0.20	0.60	1.00
40	0.21	0.61	1.00

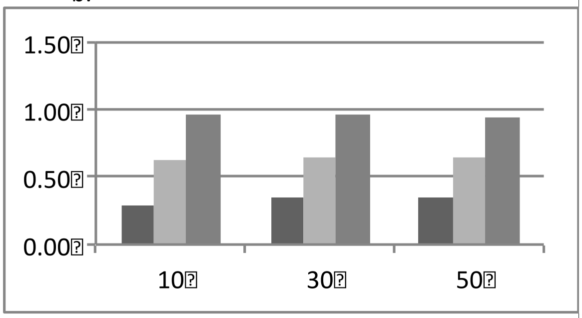
a.



Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.28	0.62	0.96
30	0.34	0.65	0.96
50	0.35	0.65	0.95

b.

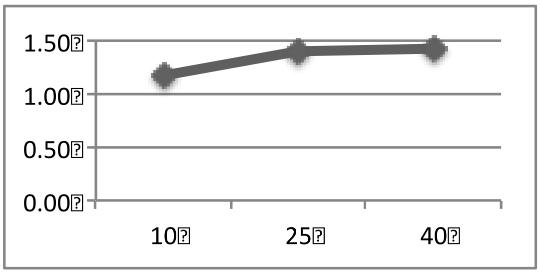


b) Clustering coefficient development for user(a) and article(b) networks when GAM = 5.

Average Small World Coefficient in User Network

Step	AVG
10	1.17
25	1.41
40	1.43

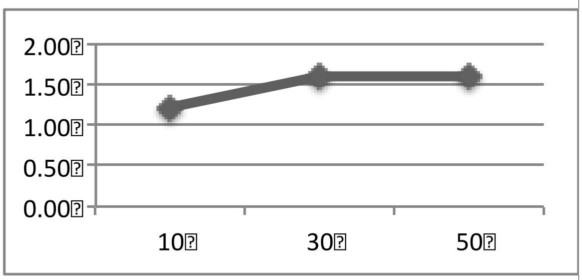
a.



Average Small World Coefficient in Article Network

Step	AVG
10	1.22
30	1.59
50	1.60

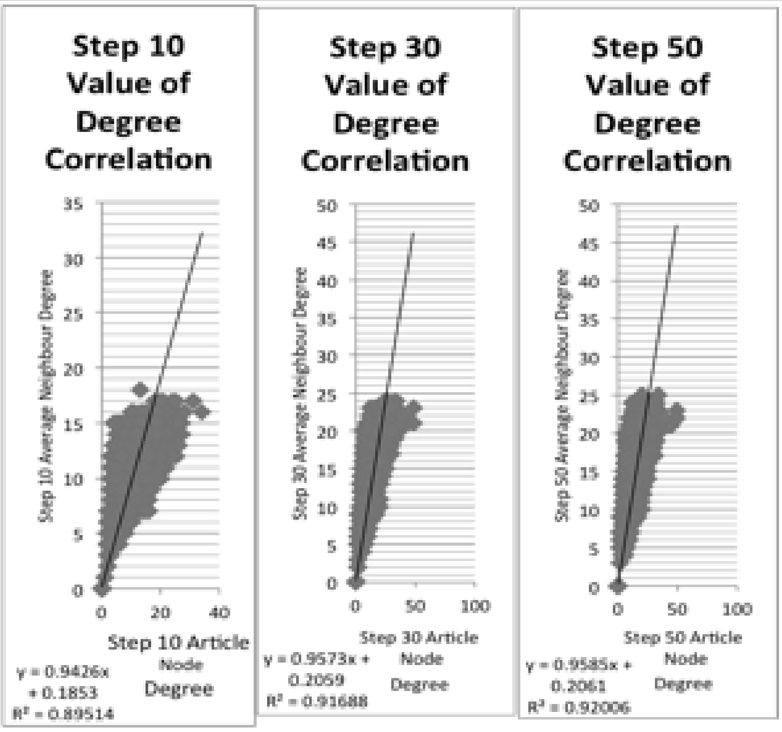
b.



c) Small World coefficient development for user(a) and article(b) networks when GAM = 5.

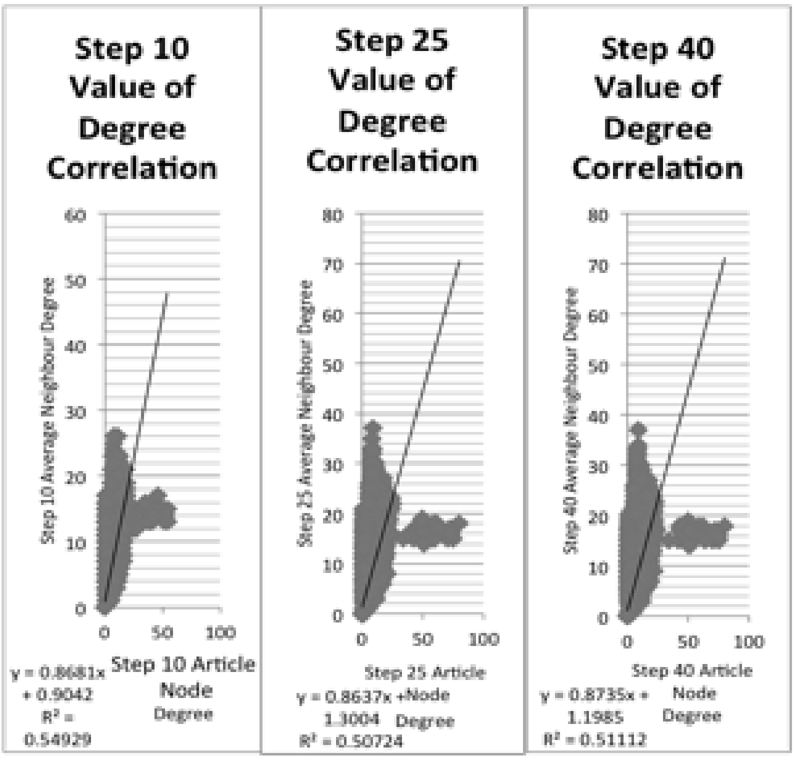
Fig. 48 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GAM=5.

Article Network Degree Correlation Values



a) Degree correlation development for article network when GAM = 5.

User Network Degree Correlation Values



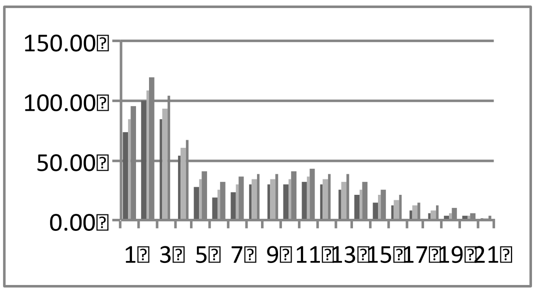
b) Degree correlation development for user network when GAM = 5.

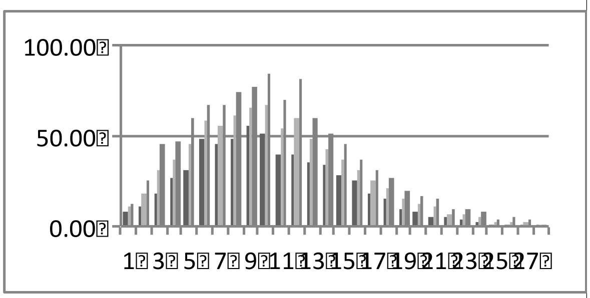
Fig .49 Degree correlation development for GAM = 5.

Table 19 Article and User Network Degree Distribution for GAM = 5

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	73.05	84.20	95.35
2	100.62	109.50	118.38
3	83.82	94.00	104.18
4	54.46	60.60	66.74
5	26.96	33.80	40.64
6	19.48	25.45	31.42
7	23.91	30.45	36.99
8	29.04	33.75	38.46
9	30.20	33.95	37.70
10	29.17	35.25	41.33
11	31.36	37.25	43.14
12	30.78	35.05	39.32
13	26.07	32.05	38.03
14	20.05	25.85	31.65
15	14.93	20.10	25.27
16	12.48	16.60	20.72
17	9.07	12.10	15.13
18	5.77	8.75	11.73
19	2.52	6.00	9.48
20	2.59	4.10	5.61
21	0.13	2.40	4.67

Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	7.87	10.50	13.13
2	10.58	18.00	25.42
3	18.30	31.75	45.20
4	26.80	36.58	46.37
5	31.07	45.17	59.26
6	48.37	58.00	67.63
7	44.94	56.00	67.06
8	49.02	61.58	74.14
9	55.80	66.25	76.70
10	51.46	67.67	83.87
11	39.06	54.17	69.28
12	39.60	60.50	81.40
13	35.36	47.92	60.48
14	33.34	42.33	51.33
15	27.87	36.33	44.79
16	25.52	31.00	36.48
17	18.84	24.92	30.99
18	16.01	21.33	26.66
19	10.25	14.83	19.41
20	8.53	12.50	16.47
21	5.56	10.58	15.61
22	4.76	7.08	9.41
23	3.52	6.58	9.65
24	2.42	4.92	7.42
25	1.06	2.75	4.44
26	0.63	2.92	5.21
27	0.31	1.83	3.35
28	0.03	0.83	1.63





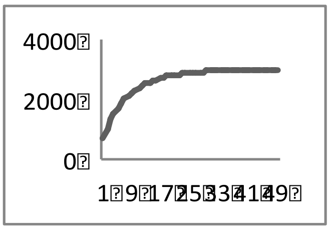
Degree distribution of User(a) and Article(b) networks when GAM = 5.

Fig .50 Degree distribution for GAM = 5.

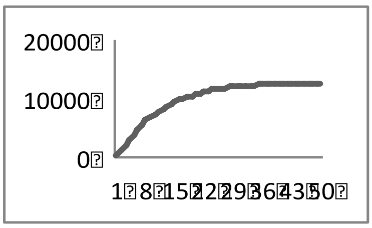
Good Article Multiplier (GAM) = 1.5, n=20 Replications Results

Table 20 Total Edge Count Development for GAM = 5

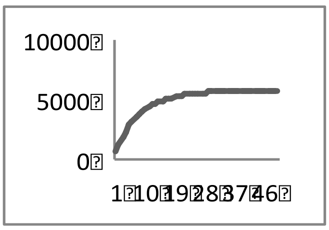
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	682.84	695.45	708.06	286.66	293.40	300.14	639.74	665.65	691.56
Step 2	1021.90	1036.25	1050.60	1016.91	1040.85	1064.79	1175.06	1237.20	1299.34
Step 3	1288.30	1305.15	1322.00	1926.76	1973.00	2019.24	1747.12	1827.25	1907.38
Step 4	1508.72	1527.30	1545.88	2891.01	2948.40	3005.79	2271.66	2369.25	2466.84
Step 5	1694.37	1714.90	1735.43	3813.00	3897.90	3982.80	2729.11	2837.50	2945.89
Step 6	1852.74	1875.40	1898.06	4682.65	4769.70	4856.75	3097.64	3221.25	3344.86
Step 7	1985.86	2010.50	2035.14	5438.76	5556.00	5673.24	3394.00	3534.60	3675.20
Step 8	2100.59	2125.20	2149.81	6109.49	6247.00	6384.51	3642.78	3794.35	3945.92
Step 9	2203.86	2228.85	2253.84	6727.48	6879.05	7030.62	3881.70	4033.35	4185.00
Step 10	2290.65	2315.55	2340.45	7261.62	7421.10	7580.58	4078.90	4230.85	4382.80
Step 11	2364.93	2392.90	2420.87	7737.52	7917.95	8098.38	4249.07	4400.85	4552.63
Step 12	2430.39	2461.10	2491.81	8155.65	8373.75	8591.85	4399.40	4549.65	4699.90
Step 13	2486.89	2519.05	2551.21	8548.52	8764.50	8980.48	4529.20	4686.35	4843.50
Step 14	2540.59	2573.55	2606.51	8910.50	9139.25	9368.00	4640.80	4803.25	4965.70
Step 15	2586.62	2621.00	2655.38	9229.35	9464.25	9699.15	4750.54	4908.75	5066.96
Step 16	2628.08	2662.50	2696.92	9533.11	9775.00	10016.89	4841.72	5006.25	5170.78
Step 17	2665.89	2701.70	2737.51	9816.11	10063.90	10311.69	4929.51	5091.45	5253.39
Step 18	2698.41	2735.10	2771.79	10076.40	10328.35	10580.30	5001.82	5170.25	5338.68
Step 19	2728.63	2766.50	2804.37	10333.49	10580.75	10828.01	5072.15	5242.05	5411.95
Step 20	2752.45	2791.60	2830.75	10526.87	10782.80	11038.73	5125.93	5301.35	5476.77
Step 21	2775.04	2815.70	2856.36	10696.48	10988.20	11279.92	5181.89	5359.75	5537.61
Step 22	2794.54	2837.15	2879.76	10880.31	11180.05	11479.79	5228.64	5411.55	5594.46
Step 23	2812.11	2855.55	2898.99	11043.05	11353.05	11663.05	5272.58	5460.95	5649.32
Step 24	2825.13	2870.55	2915.97	11170.83	11488.95	11807.07	5302.32	5493.25	5684.18
Step 25	2837.83	2885.50	2933.17	11293.09	11634.10	11975.11	5336.77	5530.45	5724.13
Step 26	2850.05	2897.50	2944.95	11384.38	11742.15	12099.92	5356.21	5552.85	5749.49
Step 27	2860.85	2908.60	2956.35	11476.66	11846.25	12215.84	5377.32	5579.25	5781.18
Step 28	2871.03	2918.65	2966.27	11582.85	11944.85	12306.85	5403.00	5603.05	5803.10
Step 29	2878.20	2926.50	2974.80	11634.80	12014.80	12394.80	5417.17	5616.65	5816.13
Step 30	2885.53	2934.05	2982.57	11716.24	12093.05	12469.86	5435.96	5632.55	5829.14
Step 31	2893.14	2940.95	2988.76	11786.51	12157.50	12528.49	5450.54	5645.85	5841.16
Step 32	2900.05	2947.90	2995.75	11835.80	12221.80	12607.80	5464.74	5658.65	5852.56
Step 33	2906.04	2953.40	3000.76	11884.52	12281.10	12677.68	5478.26	5671.05	5863.84
Step 34	2911.38	2958.55	3005.72	11931.47	12332.40	12733.33	5493.03	5681.35	5869.67
Step 35	2914.89	2962.45	3010.01	11972.00	12371.60	12771.20	5505.64	5690.05	5874.46
Step 36	2918.78	2966.25	3013.72	12002.86	12410.50	12818.14	5512.77	5696.65	5880.53
Step 37	2922.41	2969.65	3016.89	12038.86	12444.60	12850.34	5521.03	5703.25	5885.47
Step 38	2925.71	2972.80	3019.89	12066.57	12477.90	12889.23	5528.46	5710.35	5892.24
Step 39	2928.40	2975.60	3022.80	12085.78	12510.70	12935.62	5535.68	5716.55	5897.42
Step 40	2930.82	2978.05	3025.28	12105.99	12531.80	12957.61	5539.30	5720.15	5901.00
Step 41	2933.17	2980.35	3027.53	12129.86	12553.90	12977.94	5542.45	5724.55	5906.65
Step 42	2934.60	2982.40	3030.20	12149.06	12575.90	13002.74	5546.14	5728.55	5910.96
Step 43	2937.03	2984.65	3032.27	12171.49	12601.05	13030.61	5552.51	5734.45	5916.39
Step 44	2938.98	2986.40	3033.82	12184.05	12617.45	13050.85	5556.53	5737.85	5919.17
Step 45	2941.52	2989.05	3036.58	12202.99	12641.25	13079.51	5560.95	5742.45	5923.95
Step 46	2942.74	2990.70	3038.66	12216.93	12661.90	13106.87	5563.87	5746.15	5928.43
Step 47	2944.32	2992.30	3040.28	12237.75	12678.50	13119.25	5565.44	5748.15	5930.86
Step 48	2945.92	2993.70	3041.48	12260.78	12693.60	13126.42	5570.90	5752.15	5933.40
Step 49	2946.92	2994.65	3042.38	12275.58	12704.80	13134.02	5572.50	5754.55	5936.60
Step 50	2947.44	2995.50	3043.56	12284.61	12716.55	13148.49	5573.39	5756.45	5939.51



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

Fig. 51 Network total edge count development for GAM = 1.5.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

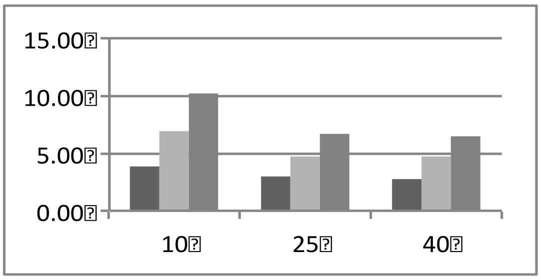
Step	AVG-STDEV	AVG	AVG+STDEV
10	3.78	7.00	10.21
25	2.91	4.83	6.76
40	2.85	4.62	6.40

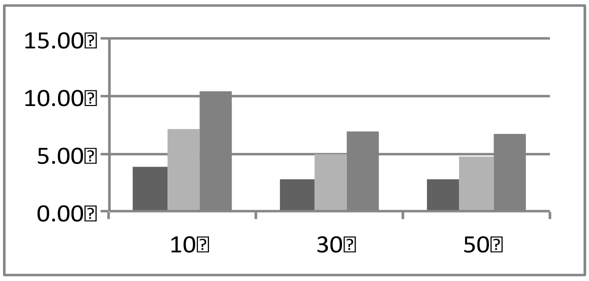
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.96	7.15	10.34
30	2.87	4.88	6.89
50	2.73	4.68	6.64

b.





a) Pathlength development for user(a) and article(b) networks when GAM = 1.5

Average Clustering Coefficient in User Network

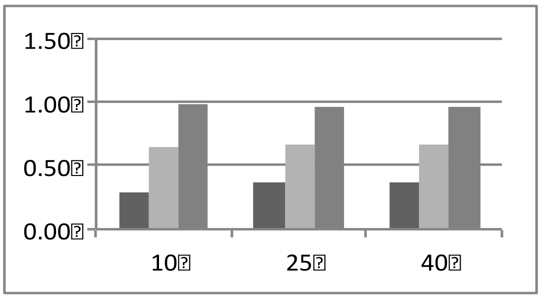
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.30	0.64	0.98
25	0.37	0.67	0.97
40	0.36	0.66	0.96

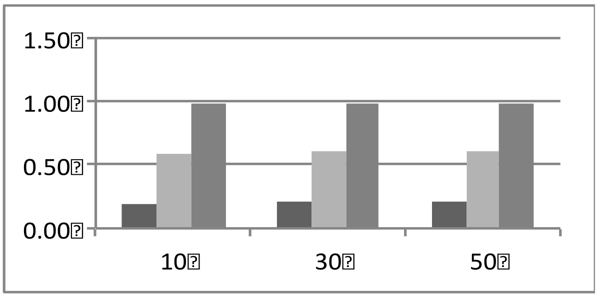
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.18	0.58	0.98
30	0.21	0.60	0.99
50	0.21	0.60	0.99

b.





b) Clustering coefficient development for user(a) and article(b) networks when GAM = 1.5

Average Small World Coefficient in User Network

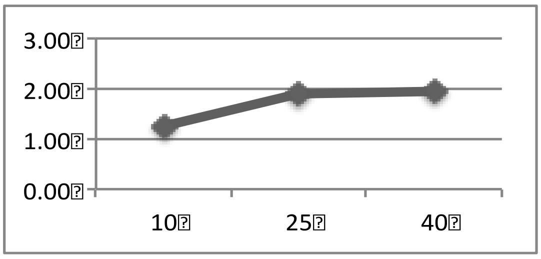
Step	AVG
10	1.26
25	1.91
40	1.97

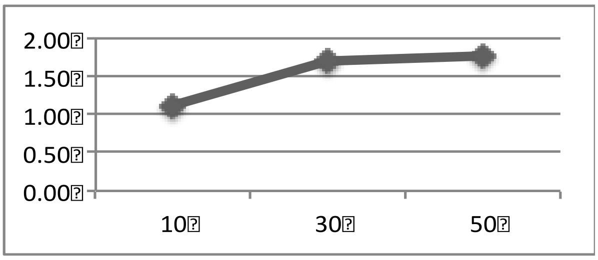
a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.12
30	1.69
50	1.76

b.

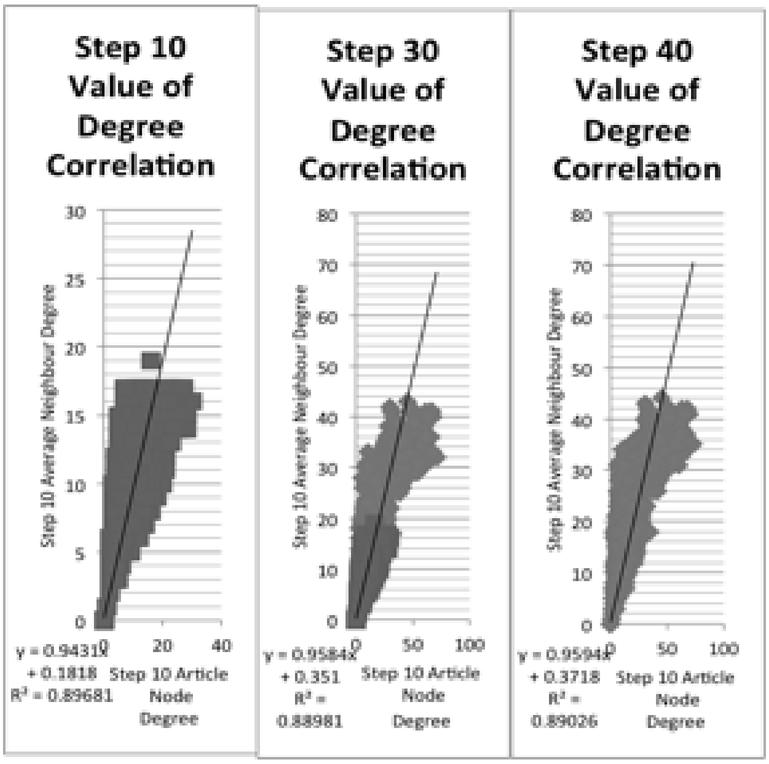




c) Small World coefficient development for user(a) and article(b) networks when GAM = 1.5

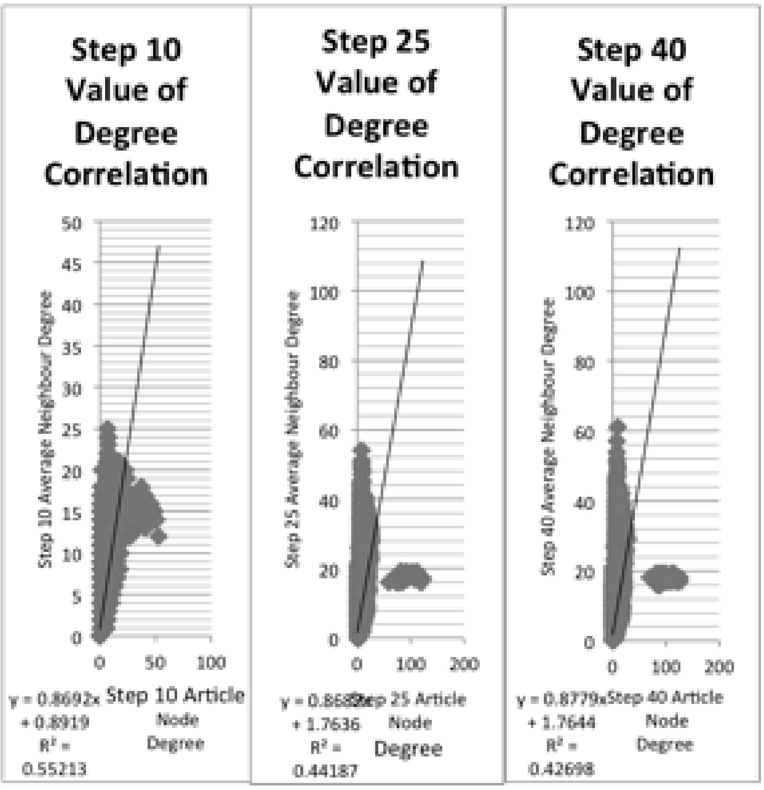
Fig. 52 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for GAM= 1.5.

Article Network Degree Correlation Values



a) Degree distribution development for article network when GAM = 1.5

User Network Degree Correlation Values

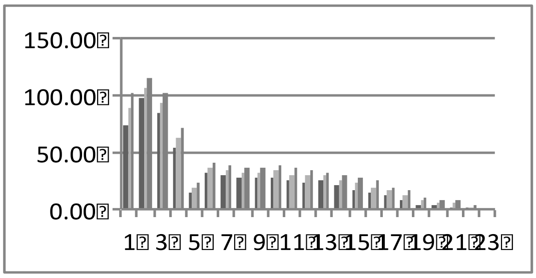


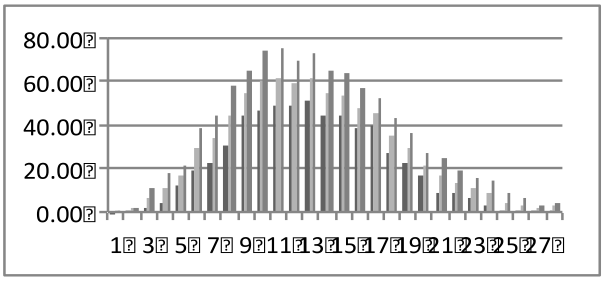
b) Degree distribution development for user network when GAM = 1.5

Fig. 53 Degree correlation development for GAM = 1.5.

Table 21 Article and User Network Degree Distribution for GAM = 1.5

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG+STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG+STDEV
1	73.17	87.85	102.53	1	-0.14	0.33	0.80
2	98.31	106.70	115.09	2	0.23	1.33	2.44
3	84.51	93.55	102.59	3	1.42	6.08	10.75
4	53.93	63.10	72.27	4	3.96	11.00	18.04
5	14.26	18.35	22.44	5	11.88	16.75	21.62
6	31.51	36.15	40.79	6	19.67	29.17	38.66
7	28.74	33.35	37.96	7	22.39	33.67	44.94
8	26.58	31.45	36.32	8	30.96	44.33	57.70
9	26.97	31.95	36.93	9	44.94	55.08	65.23
10	27.79	33.65	39.51	10	46.27	60.00	73.73
11	24.96	30.45	35.94	11	48.97	62.08	75.19
12	23.86	28.90	33.94	12	48.82	59.25	69.68
13	24.59	28.75	32.91	13	50.89	61.92	72.94
14	20.87	25.60	30.33	14	44.22	54.50	64.78
15	16.41	22.20	27.99	15	44.21	53.83	63.46
16	14.87	19.70	24.53	16	38.40	47.92	57.43
17	12.32	16.05	19.78	17	39.73	46.08	52.44
18	8.78	13.10	17.42	18	27.68	35.58	43.48
19	4.54	7.55	10.56	19	22.02	29.00	35.98
20	3.63	6.10	8.57	20	16.78	21.75	26.72
21	2.51	4.95	7.39	21	8.84	16.67	24.49
22	0.33	2.05	3.77	22	8.40	13.67	18.93
				23	6.84	11.42	15.99
				24	3.12	8.67	14.21
				25	0.85	4.75	8.65
				26	0.77	3.58	6.40
				27	0.90	2.08	3.27
				28	0.61	2.50	4.39





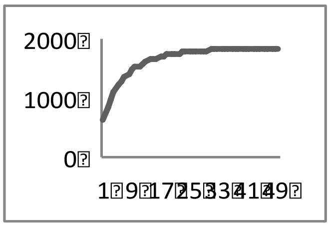
Degree distribution of User(a) and Article(b) networks when GAM = 1.5.

Fig. 54 Degree distribution for GAM = 1.5

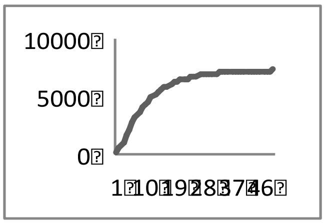
Active User Percentage (AUP) = 0.2, n=20 Replications Results

Table 22 Total Edge Count Development for AUP = 0.2

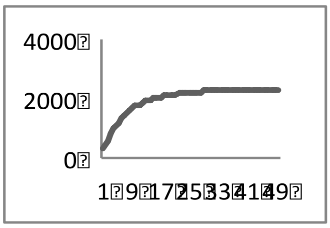
	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	614.94	623.75	632.56	155.45	160.35	165.25	304.91	319.15	333.39
Step 2	824.97	837.85	850.73	540.27	565.40	590.53	536.94	563.65	590.36
Step 3	969.58	986.05	1002.52	1021.37	1064.50	1107.63	751.35	792.85	834.35
Step 4	1088.61	1107.85	1127.09	1551.19	1620.90	1690.61	953.90	1007.05	1060.20
Step 5	1186.83	1207.40	1227.97	2081.73	2166.50	2251.27	1128.33	1189.35	1250.37
Step 6	1264.56	1289.55	1314.54	2567.61	2677.25	2786.89	1272.49	1338.90	1405.31
Step 7	1331.92	1358.80	1385.68	2995.62	3145.05	3294.48	1398.05	1467.40	1536.75
Step 8	1390.19	1418.55	1446.91	3427.23	3574.65	3722.07	1503.02	1574.30	1645.58
Step 9	1441.27	1470.25	1499.23	3830.61	3978.15	4125.69	1595.45	1667.85	1740.25
Step 10	1482.22	1513.95	1545.68	4156.35	4329.30	4502.25	1670.47	1746.55	1822.63
Step 11	1517.46	1549.60	1581.74	4441.41	4623.50	4805.59	1732.88	1807.95	1883.02
Step 12	1548.64	1581.45	1614.26	4708.37	4898.80	5089.23	1783.08	1861.75	1940.42
Step 13	1577.77	1610.55	1643.33	4951.00	5149.30	5347.60	1838.95	1913.50	1988.05
Step 14	1603.21	1636.35	1669.49	5165.92	5375.90	5585.88	1882.99	1958.30	2033.61
Step 15	1624.94	1659.10	1693.26	5362.25	5582.50	5802.75	1921.31	2000.55	2079.79
Step 16	1641.84	1678.45	1715.06	5523.61	5760.45	5997.29	1949.21	2034.40	2119.59
Step 17	1658.11	1695.45	1732.79	5684.87	5923.50	6162.13	1977.66	2064.15	2150.64
Step 18	1672.94	1710.70	1748.46	5821.70	6060.25	6298.80	2000.72	2088.75	2176.78
Step 19	1686.39	1724.75	1763.11	5944.90	6189.70	6434.50	2025.51	2114.00	2202.49
Step 20	1696.91	1737.05	1777.19	6047.12	6303.50	6559.88	2040.64	2133.60	2226.56
Step 21	1707.46	1748.00	1788.54	6137.68	6401.35	6665.02	2064.62	2155.70	2246.78
Step 22	1716.00	1756.90	1797.80	6222.10	6484.95	6747.80	2078.93	2170.80	2262.67
Step 23	1724.10	1765.60	1807.10	6291.33	6568.50	6845.67	2095.49	2186.55	2277.61
Step 24	1732.10	1773.30	1814.50	6371.48	6646.20	6920.92	2108.67	2199.75	2290.83
Step 25	1737.90	1780.70	1823.50	6417.37	6713.05	7008.73	2118.45	2211.05	2303.65
Step 26	1744.12	1786.70	1829.28	6482.27	6774.90	7067.53	2130.26	2221.90	2313.54
Step 27	1749.64	1791.95	1834.26	6539.15	6832.75	7126.35	2139.50	2229.80	2320.10
Step 28	1754.48	1796.85	1839.22	6582.99	6878.80	7174.61	2147.37	2237.70	2328.03
Step 29	1757.94	1800.60	1843.26	6617.88	6914.95	7212.02	2153.73	2245.00	2336.27
Step 30	1762.34	1805.20	1848.06	6660.35	6960.00	7259.65	2162.96	2254.20	2345.44
Step 31	1765.84	1808.95	1852.06	6697.56	6997.35	7297.14	2167.87	2259.50	2351.13
Step 32	1769.24	1812.45	1855.66	6729.62	7030.65	7331.68	2172.23	2265.20	2358.17
Step 33	1771.88	1815.35	1858.82	6754.89	7059.80	7364.71	2177.09	2271.90	2366.71
Step 34	1774.40	1817.75	1861.10	6785.73	7084.45	7383.17	2181.90	2275.80	2369.70
Step 35	1776.94	1819.60	1862.26	6810.72	7104.65	7398.58	2186.03	2278.40	2370.77
Step 36	1778.25	1821.30	1864.35	6832.53	7125.50	7418.47	2187.36	2280.40	2373.44
Step 37	1779.76	1823.25	1866.74	6844.70	7144.35	7444.00	2189.03	2283.90	2378.77
Step 38	1780.90	1824.50	1868.10	6852.45	7157.10	7461.75	2190.31	2285.40	2380.49
Step 39	1782.36	1826.00	1869.64	6864.71	7171.70	7478.69	2192.81	2288.00	2383.19
Step 40	1783.56	1827.20	1870.84	6876.39	7185.35	7494.31	2193.99	2289.60	2385.21
Step 41	1784.62	1828.40	1872.18	6888.14	7197.95	7507.76	2195.34	2291.30	2387.26
Step 42	1785.25	1829.10	1872.95	6893.29	7204.40	7515.51	2196.53	2292.40	2388.27
Step 43	1785.52	1829.85	1874.18	6894.26	7212.25	7530.24	2196.80	2293.10	2389.40
Step 44	1786.56	1830.65	1874.74	6905.58	7220.45	7535.32	2198.07	2294.40	2390.73
Step 45	1787.31	1831.50	1875.69	6913.13	7231.65	7550.17	2199.77	2295.70	2391.63
Step 46	1788.10	1832.10	1876.10	6922.20	7238.35	7554.50	2201.26	2296.80	2392.34
Step 47	1788.54	1832.90	1877.26	6926.46	7245.35	7564.24	2202.01	2297.90	2393.79
Step 48	1788.91	1833.30	1877.69	6932.51	7251.55	7570.59	2202.63	2298.50	2394.37
Step 49	1789.65	1834.00	1878.35	6937.68	7257.05	7576.42	2203.74	2299.60	2395.46
Step 50	1789.97	1834.75	1879.53	6941.81	7265.15	7588.49	2204.24	2300.60	2396.96



Affiliation Network Total Edge Count



Article Network Total Edge Count



User Network Total Edge Count

Fig. 55 Network total edge count development for AUP + 0.2.

Key Network Metrics for Article and User Networks

Average pathlength in User Network

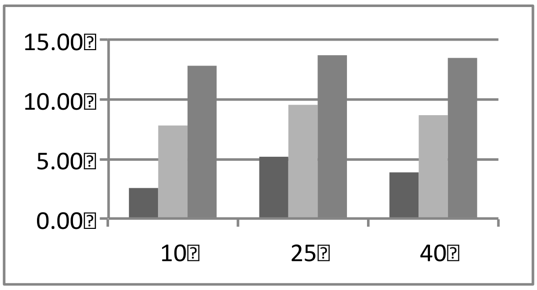
Step	AVG-STDEV	AVG	AVG+STDEV
10	2.52	7.70	12.88
25	5.18	9.44	13.70
40	3.94	8.75	13.55

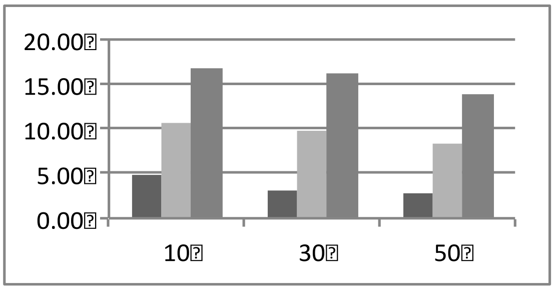
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	4.90	10.82	16.74
30	3.19	9.71	16.22
50	2.83	8.41	14.00

b.





a) Pathlength development for user(a) and article(b) networks when AUP = 0.2

Average Clustering Coefficient in User Network

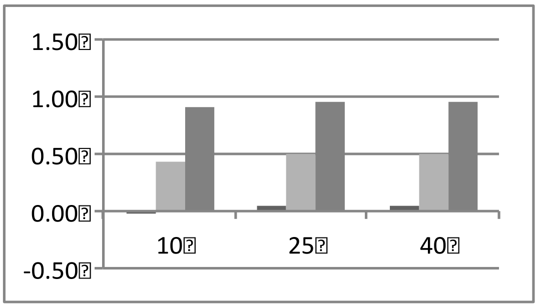
Step	AVG-STDEV	AVG	AVG+STDEV
10	-0.02	0.44	0.90
25	0.04	0.50	0.96
40	0.04	0.50	0.96

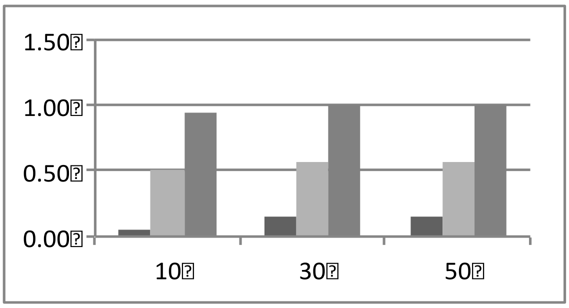
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.05	0.50	0.95
30	0.15	0.57	1.00
50	0.15	0.58	1.00

b.





b) Clustering coefficient development for user(a) and article(b) networks when AUP = 0.2

Average Small World Coefficient in User Network

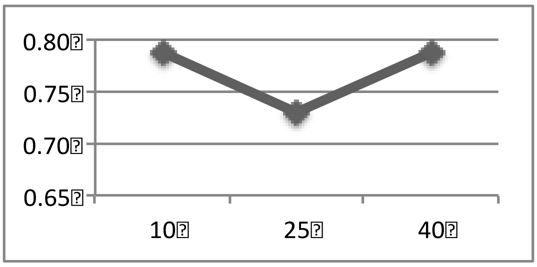
Step	AVG
10	0.79
25	0.73
40	0.79

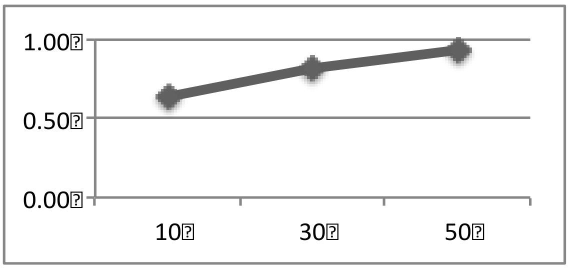
a.

Average Small World Coefficient in Article Network

Step	AVG
10	0.64
30	0.81
50	0.94

b.

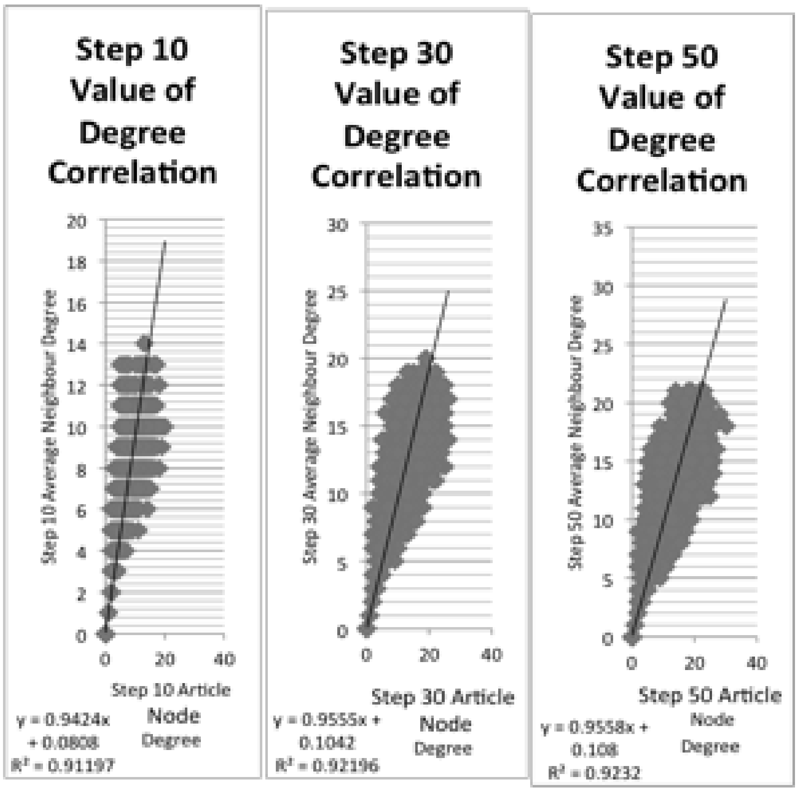




c) Small World coefficient development for user(a) and article(b) networks when AUP = 0.2

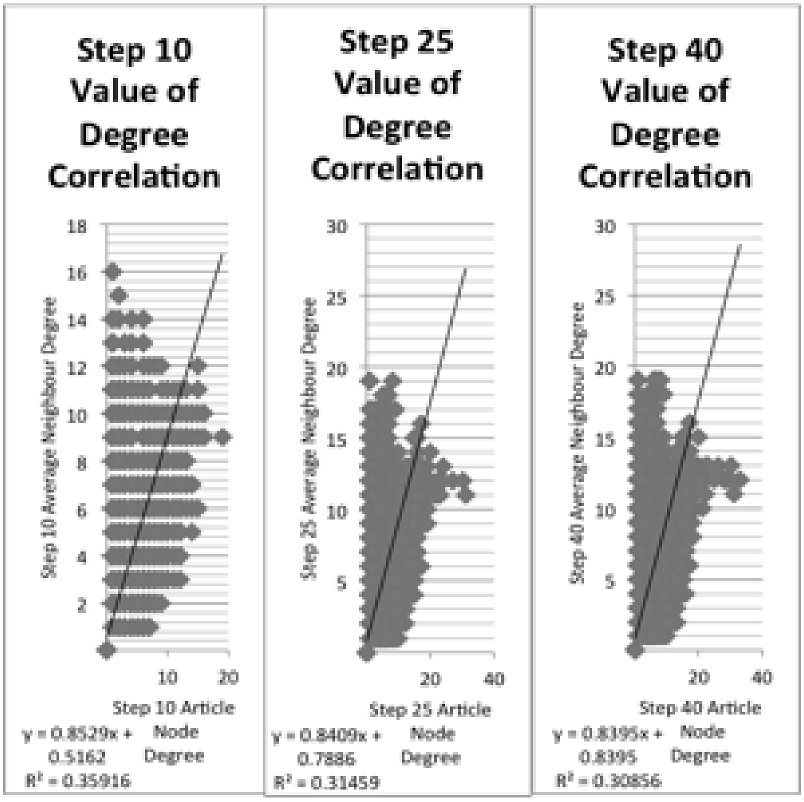
Fig. 56 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. for AUP= 0.2.

Article Network Degree Correlation Values



a) Degree distribution development for article network when AUP = 0.2

User Network Degree Correlation Values

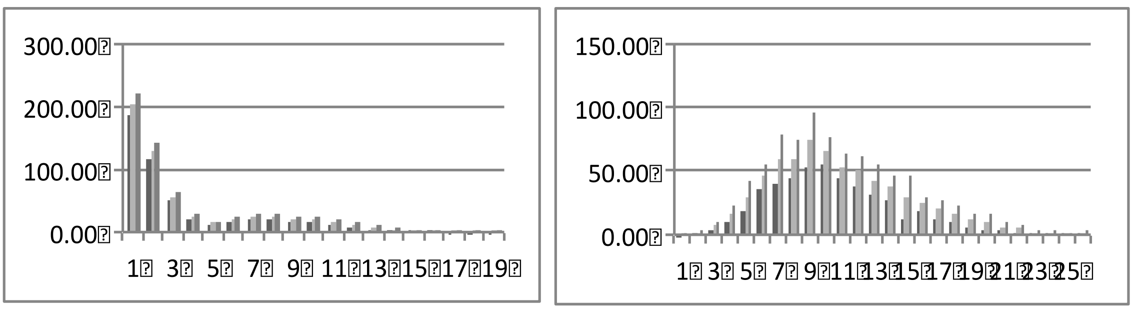


b) Degree distribution development for user network when AUP = 0.2

Fig. 57 Degree correlation development for AUP = 0.2.

Table 23 Article and User Network Degree Distribution for AUP = 0.2

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG=STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG=STDEV
1	185.34	203.50	221.66	1	-0.22	0.42	1.06
2	115.33	130.05	144.77	2	0.77	2.25	3.73
3	50.51	56.75	62.99	3	3.26	7.00	10.74
4	20.04	24.40	28.76	4	10.27	16.67	23.06
5	9.85	13.95	18.05	5	18.26	30.08	41.90
6	17.25	21.25	25.25	6	35.61	45.92	56.22
7	18.91	23.70	28.49	7	39.99	59.17	78.34
8	19.00	24.15	29.30	8	45.44	60.08	74.73
9	15.15	20.75	26.35	9	53.93	74.92	95.90
10	16.09	19.80	23.51	10	54.99	66.17	77.35
11	13.20	17.10	21.00	11	43.67	53.33	63.00
12	7.88	11.70	15.52	12	38.27	50.00	61.73
13	3.14	6.45	9.76	13	31.47	42.92	54.37
14	2.40	4.30	6.20	14	27.25	37.25	47.25
15	1.89	3.35	4.81	15	11.32	29.17	47.01
16	0.27	2.00	3.73	16	18.71	24.42	30.12
17	-0.33	0.95	2.23	17	12.92	20.00	27.08
18	-0.41	0.65	1.71	18	9.27	16.00	22.73
19	-0.22	0.35	0.92	19	5.88	11.25	16.62
				20	3.59	9.67	15.74
				21	2.54	5.92	9.29
				22	1.93	4.92	7.90
				23	0.13	2.00	3.87
				24	0.15	2.08	4.02
				25	0.09	1.00	1.91
				26	0.03	1.42	2.80



Degree distribution of User(a) and Article(b) networks.

Fig. 58 Degree distribution for AUP = 0.2.

Active User Percentage (AUP) = 0.6, n=20 Replication Results

Table 24 Total Edge Count Development for AUP = 0.6

	Affiliation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	749.32	761.47	773.63	390.32	398.00	405.68	1003.73	1047.47	1091.22
Step 2	1218.19	1234.89	1251.60	1410.19	1435.21	1460.23	1949.71	2050.47	2151.24
Step 3	1600.13	1622.26	1644.40	2653.57	2722.32	2791.06	2995.05	3137.58	3280.11
Step 4	1923.50	1948.05	1972.61	3940.83	4045.26	4149.70	3967.41	4128.58	4289.74
Step 5	2189.24	2219.74	2250.23	5119.88	5266.11	5412.33	4760.50	4935.32	5110.13
Step 6	2420.89	2455.37	2489.85	6208.98	6383.74	6558.49	5439.32	5617.53	5795.73
Step 7	2616.62	2655.42	2694.22	7167.22	7351.79	7536.36	5970.19	6182.05	6393.92
Step 8	2790.45	2828.37	2866.29	8041.49	8235.84	8430.19	6413.30	6631.32	6849.33
Step 9	2937.17	2978.84	3020.51	8780.72	9014.11	9247.49	6793.79	7017.74	7241.68
Step 10	3062.14	3108.53	3154.91	9404.82	9676.79	9948.76	7106.36	7344.05	7581.75
Step 11	3177.16	3223.26	3269.37	10011.72	10294.11	10576.49	7383.08	7621.42	7859.77
Step 12	3280.00	3325.32	3370.64	10573.44	10853.58	11133.72	7633.86	7877.63	8121.41
Step 13	3369.17	3414.42	3459.67	11092.09	11359.79	11627.49	7861.61	8100.68	8339.76
Step 14	3447.84	3492.74	3537.63	11549.50	11828.05	12106.60	8063.76	8298.68	8533.61
Step 15	3510.25	3558.37	3606.49	11947.70	12243.42	12539.15	8223.54	8470.58	8717.62
Step 16	3571.61	3622.42	3673.23	12336.13	12648.42	12960.71	8389.02	8634.37	8879.72
Step 17	3624.59	3676.42	3728.25	12693.73	13021.16	13348.58	8532.75	8776.37	9019.99
Step 18	3671.47	3724.58	3777.69	13031.29	13374.63	13717.98	8664.26	8906.89	9149.53
Step 19	3714.82	3769.16	3823.50	13354.27	13712.32	14070.36	8779.56	9020.68	9261.81
Step 20	3751.04	3808.21	3865.38	13669.39	14041.16	14412.92	8879.00	9123.42	9367.85
Step 21	3785.52	3844.32	3903.11	13970.56	14357.16	14743.76	8978.16	9225.84	9473.53
Step 22	3816.72	3877.79	3938.86	14263.56	14670.00	15076.44	9067.82	9327.95	9588.07
Step 23	3843.60	3906.00	3968.40	14541.71	14962.26	15382.82	9155.94	9424.26	9692.59
Step 24	3869.96	3932.42	3994.88	14832.86	15256.58	15680.30	9231.82	9505.84	9779.87
Step 25	3889.36	3954.53	4019.70	15103.57	15529.16	15954.74	9304.73	9579.32	9853.90
Step 26	3911.83	3977.21	4042.59	15405.16	15835.79	16266.42	9378.71	9657.84	9936.97
Step 27	3932.46	3999.05	4065.65	15709.42	16141.79	16574.16	9457.78	9734.05	10010.32
Step 28	3950.86	4018.05	4085.25	15999.59	16445.68	16891.78	9526.38	9800.68	10074.99
Step 29	3966.74	4035.26	4103.78	16283.32	16752.21	17221.10	9583.71	9859.84	10135.98
Step 30	3980.93	4050.63	4120.33	16580.92	17054.84	17528.76	9643.22	9920.79	10198.36
Step 31	3994.47	4065.05	4135.64	16857.48	17345.63	17833.79	9694.19	9972.79	10251.39
Step 32	4007.01	4078.32	4149.62	17130.21	17633.26	18136.32	9742.18	10022.37	10302.56
Step 33	4017.18	4090.05	4162.93	17394.07	17914.84	18435.61	9788.64	10072.05	10355.46
Step 34	4029.38	4102.53	4175.67	17721.40	18233.42	18745.45	9846.89	10128.37	10409.85
Step 35	4041.02	4113.42	4185.82	18010.49	18524.47	19038.46	9893.25	10177.21	10461.18
Step 36	4051.83	4124.42	4197.01	18309.86	18832.53	19355.19	9942.66	10226.68	10510.71
Step 37	4063.16	4135.16	4207.16	18599.72	19144.58	19689.44	9982.43	10269.63	10556.83
Step 38	4072.29	4144.95	4217.61	18897.74	19465.05	20032.36	10032.71	10319.53	10606.34
Step 39	4081.33	4154.42	4227.51	19182.46	19774.89	20367.33	10080.30	10363.95	10647.59
Step 40	4088.29	4162.47	4236.66	19409.88	20057.58	20705.28	10111.20	10403.42	10695.64
Step 41	4096.01	4170.00	4243.99	19613.91	20306.74	20999.56	10139.32	10435.32	10731.31
Step 42	4101.81	4177.00	4252.19	19817.48	20557.68	21297.88	10163.45	10464.68	10765.92
Step 43	4106.06	4183.37	4260.68	19907.96	20763.32	21618.67	10181.98	10491.84	10801.70
Step 44	4108.77	4187.89	4267.02	19952.52	20913.21	21873.90	10195.00	10512.05	10829.11
Step 45	4110.56	4191.79	4273.02	19980.48	21040.74	22100.99	10200.85	10528.05	10855.26
Step 46	4112.54	4194.16	4275.78	19995.08	21097.37	22199.66	10202.39	10536.68	10870.98
Step 47	4114.30	4196.47	4278.65	20015.96	21151.37	22286.78	10203.03	10541.53	10880.03
Step 48	4115.82	4198.37	4280.92	20021.00	21182.21	22343.42	10207.29	10545.53	10883.76
Step 49	4117.63	4200.26	4282.89	20058.80	21221.26	22383.72	10210.33	10550.05	10889.77
Step 50	4118.90	4201.53	4284.16	20076.20	21239.26	22402.33	10212.79	10552.89	10893.00

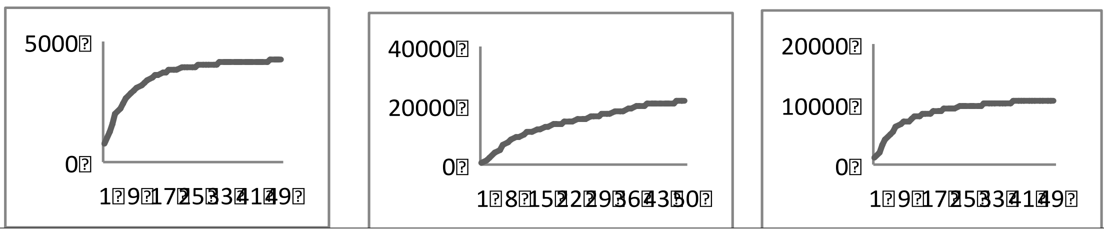


Fig. 59 Network total edge count development for AUP + 0.6.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

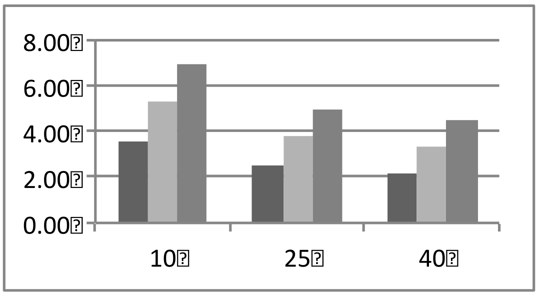
Step	AVG-STDEV	AVG	AVG+STDEV
10	3.56	5.29	7.01
25	2.52	3.77	5.03
40	2.19	3.32	4.46

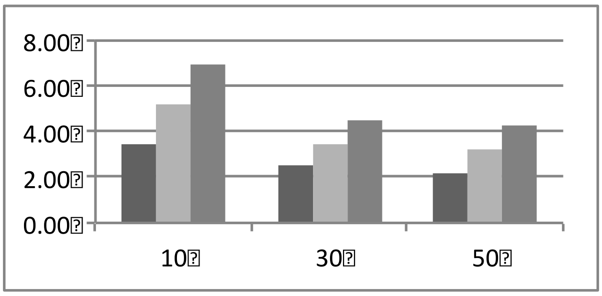
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.47	5.22	6.97
30	2.50	3.48	4.45
50	2.19	3.23	4.27

b.





a) Pathlength development for user(a) and article(b) networks when AUP = 0.6

Average Clustering Coefficient in User Network

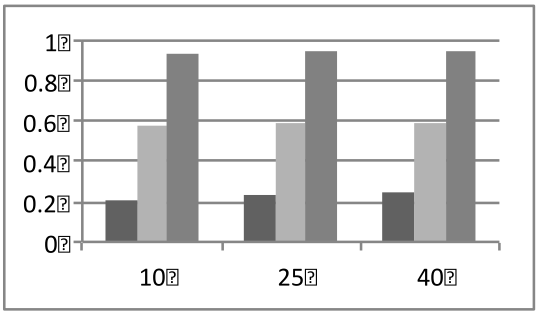
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.2134	0.5741	0.9349
25	0.2359	0.5901	0.9444
40	0.24	0.5941	0.9481

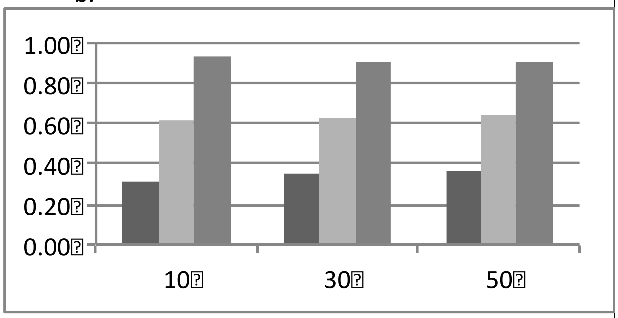
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.31	0.62	0.94
30	0.36	0.63	0.91
50	0.36	0.64	0.91

b.





b) Clustering coefficient development for user(a) and article(b) networks when AUP = 0.6

Average Small World Coefficient in User Network

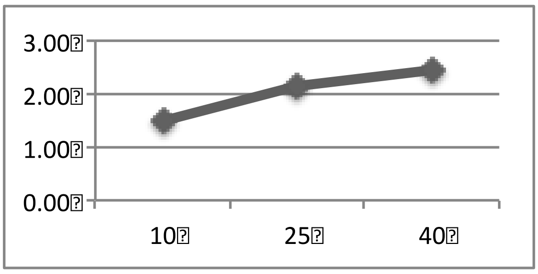
Step	AVG
10	1.49
25	2.15
40	2.46

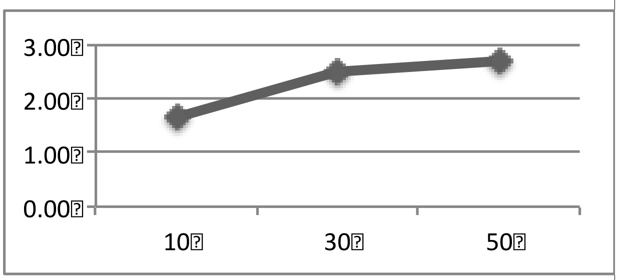
a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.64
30	2.50
50	2.71

b.

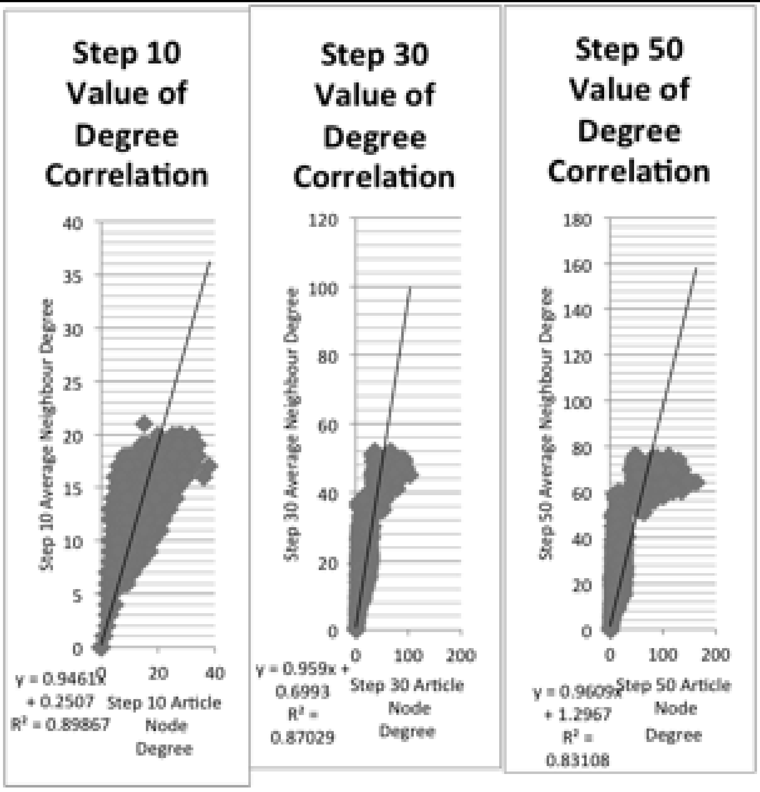




c) Small World coefficient development for user(a) and article(b) networks when AUP = 0.6

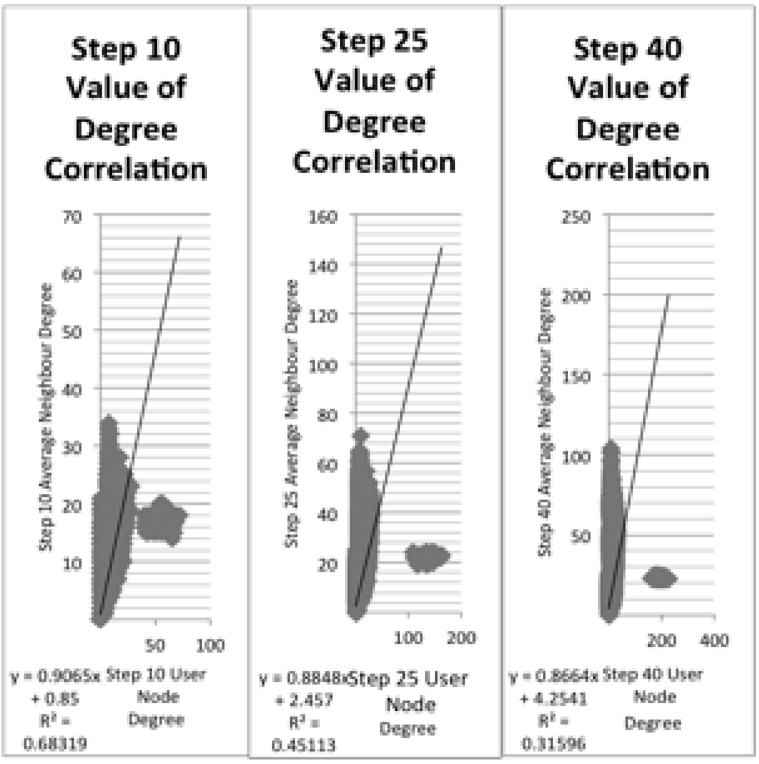
Fig. 60 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. AUP=0.6.

Article Network Degree Correlation Values



a) Degree distribution development for article network when AUP = 0.6

User Network Degree Correlation Values



b) Degree distribution development for user network when AUP = 0.6

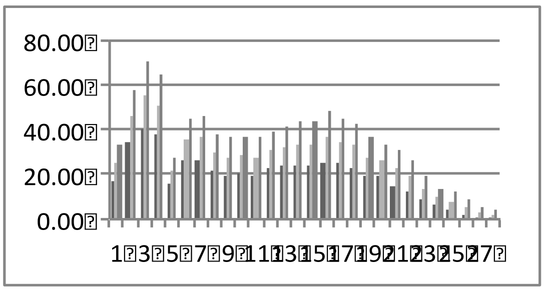
Fig. 61 Degree correlation development for AUP = 0.6.

Table 25 Article and User Network Degree Distribution for AUP = 0.6

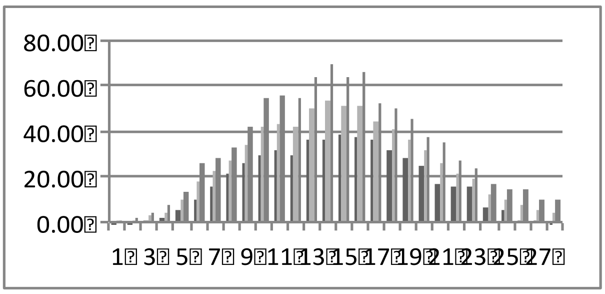
Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	16.99	25.10	33.21
2	34.98	46.35	57.72
3	40.09	55.20	70.31
4	37.57	51.45	65.33
5	15.62	21.55	27.48
6	26.15	35.40	44.65
7	26.72	36.70	46.68
8	22.31	30.45	38.59
9	19.50	28.05	36.60
10	21.09	29.15	37.21
11	18.83	27.85	36.87
12	22.54	30.95	39.36
13	23.99	32.65	41.31
14	24.19	33.80	43.41
15	23.51	33.55	43.59
16	25.51	37.00	48.49
17	24.92	34.80	44.68
18	23.34	33.05	42.76
19	19.67	28.10	36.53
20	18.86	25.90	32.94
21	15.05	22.85	30.65
22	12.89	19.75	26.61
23	8.83	14.05	19.27
24	6.39	10.10	13.81
25	3.86	7.85	11.84
26	2.48	5.55	8.62
27	0.74	3.15	5.56
28	0.08	2.10	4.12

Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	-0.21	0.17	0.54
2	-0.08	0.67	1.41
3	1.02	2.83	4.65
4	1.77	4.75	7.73
5	5.27	9.58	13.89
6	10.12	18.08	26.05
7	16.02	22.42	28.81
8	21.93	27.17	32.41
9	25.82	34.17	42.51
10	29.19	41.92	54.64
11	31.47	43.75	56.03
12	29.56	41.92	54.27
13	36.91	50.50	64.09
14	36.89	53.25	69.61
15	38.97	51.25	63.53
16	37.71	51.75	65.79
17	35.84	44.42	52.99
18	31.31	40.58	49.85
19	28.19	36.75	45.31
20	25.01	31.42	37.83
21	16.93	25.92	34.90
22	16.18	21.50	26.82
23	15.21	19.67	24.13
24	6.61	11.83	17.06
25	5.88	10.00	14.12
26	1.14	8.00	14.86
27	0.50	5.17	9.84
28	-2.44	4.00	10.44

a.



b.



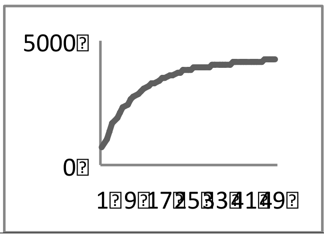
Degree distribution of User(a) and Article(b) networks when AUP = 0.6.

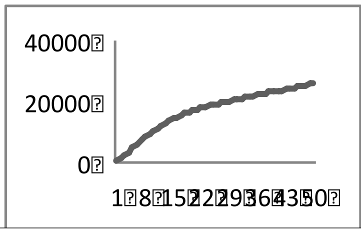
Fig. 62 Degree distribution for AUP =0.6.

Neighbourhood Dimension (ND) = 6, n=20 Replications Results

Table 26 Total Edge Count Development for ND = 6

	Affiliation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	683.46	695.70	707.94	286.41	294.20	301.99	639.81	669.00	698.19
Step 2	1043.95	1061.90	1079.85	1065.04	1088.70	1112.36	1235.81	1292.45	1349.09
Step 3	1345.37	1367.80	1390.23	2149.30	2201.20	2253.10	1944.14	2028.80	2113.46
Step 4	1602.82	1632.70	1662.58	3348.20	3450.15	3552.10	2648.06	2759.80	2871.54
Step 5	1831.22	1869.85	1908.48	4590.85	4763.40	4935.95	3309.51	3445.00	3580.49
Step 6	2043.76	2083.50	2123.24	5906.93	6084.45	6261.97	3917.58	4042.10	4166.62
Step 7	2234.14	2272.60	2311.06	7135.42	7310.05	7484.68	4416.10	4553.90	4691.70
Step 8	2401.36	2441.55	2481.74	8254.94	8458.85	8662.76	4853.86	5000.90	5147.94
Step 9	2547.59	2592.25	2636.91	9267.80	9509.70	9751.60	5219.59	5378.20	5536.81
Step 10	2680.35	2728.90	2777.45	10215.12	10488.95	10762.78	5544.95	5706.70	5868.45
Step 11	2803.71	2853.45	2903.19	11093.24	11385.25	11677.26	5832.84	6001.80	6170.76
Step 12	2913.72	2966.40	3019.08	11906.43	12215.30	12524.17	6081.24	6268.30	6455.36
Step 13	3019.16	3072.50	3125.84	12649.62	12969.40	13289.18	6306.83	6506.20	6705.57
Step 14	3114.18	3168.80	3223.42	13345.11	13674.50	14003.89	6517.22	6719.00	6920.78
Step 15	3198.18	3257.10	3316.02	13974.31	14346.40	14718.49	6699.90	6912.90	7125.90
Step 16	3273.06	3335.50	3397.94	14568.40	14944.25	15320.10	6869.17	7087.70	7306.23
Step 17	3345.18	3406.90	3468.62	15096.91	15496.95	15896.99	7023.71	7240.90	7458.09
Step 18	3406.85	3474.10	3541.35	15588.39	16016.80	16445.21	7166.18	7385.20	7604.22
Step 19	3467.35	3535.00	3602.65	16075.12	16511.80	16948.48	7296.67	7520.40	7744.13
Step 20	3522.00	3592.45	3662.90	16533.55	16984.50	17435.45	7424.55	7652.80	7881.05
Step 21	3574.33	3644.70	3715.07	16966.91	17402.60	17838.29	7529.82	7758.80	7987.78
Step 22	3618.61	3691.40	3764.19	17374.05	17817.50	18260.95	7618.19	7864.10	8110.01
Step 23	3662.82	3735.80	3808.78	17766.19	18213.70	18661.21	7721.30	7966.10	8210.90
Step 24	3699.81	3774.15	3848.49	18104.92	18573.15	19041.38	7816.06	8059.30	8302.54
Step 25	3734.51	3810.70	3886.89	18438.95	18927.80	19416.65	7902.54	8144.80	8387.06
Step 26	3768.57	3847.35	3926.13	18775.41	19276.55	19777.69	7976.83	8222.70	8468.57
Step 27	3800.40	3879.15	3957.90	19079.84	19579.15	20078.46	8045.56	8293.30	8541.04
Step 28	3827.87	3907.70	3987.53	19384.23	19882.85	20381.47	8115.09	8361.60	8608.11
Step 29	3851.34	3932.80	4014.26	19648.69	20171.25	20693.81	8179.20	8423.70	8668.20
Step 30	3876.40	3958.20	4040.00	19932.57	20470.85	21009.13	8235.54	8484.50	8733.46
Step 31	3899.75	3981.30	4062.85	20240.09	20782.20	21324.31	8292.00	8542.60	8793.20
Step 32	3919.97	4001.65	4083.33	20518.12	21067.80	21617.48	8344.02	8597.90	8851.78
Step 33	3937.53	4019.35	4101.17	20759.12	21322.75	21886.38	8389.21	8643.30	8897.39
Step 34	3955.67	4037.40	4119.13	21011.99	21589.80	22167.61	8432.73	8687.90	8943.07
Step 35	3974.24	4055.40	4136.56	21293.84	21881.25	22468.66	8481.87	8736.90	8991.93
Step 36	3988.98	4070.80	4152.62	21560.28	22145.90	22731.52	8526.12	8781.40	9036.68
Step 37	4005.24	4086.30	4167.36	21838.58	22414.50	22990.42	8566.13	8820.60	9075.07
Step 38	4020.74	4101.90	4183.06	22122.21	22693.15	23264.09	8615.15	8867.50	9119.85
Step 39	4032.33	4115.00	4197.67	22372.84	22953.15	23533.46	8647.03	8904.60	9162.17
Step 40	4043.70	4127.50	4211.30	22649.92	23238.40	23826.88	8683.51	8942.10	9200.69
Step 41	4056.26	4140.50	4224.74	22920.22	23506.75	24093.28	8719.36	8977.00	9234.64
Step 42	4068.55	4152.00	4235.45	23196.00	23782.45	24368.90	8752.35	9009.10	9265.85
Step 43	4078.13	4162.50	4246.87	23449.31	24047.05	24644.79	8786.36	9043.10	9299.84
Step 44	4087.89	4172.90	4257.91	23722.99	24338.20	24953.41	8813.55	9076.50	9339.45
Step 45	4097.11	4182.55	4267.99	23988.26	24616.80	25245.34	8845.74	9108.40	9371.06
Step 46	4107.27	4192.65	4278.03	24248.41	24880.00	25511.59	8873.74	9137.40	9401.06
Step 47	4115.55	4200.90	4286.25	24511.39	25150.30	25789.21	8896.62	9163.70	9430.78
Step 48	4124.70	4209.55	4294.40	24791.78	25443.30	26094.82	8925.93	9192.10	9458.27
Step 49	4132.64	4217.20	4301.76	25079.01	25718.70	26358.39	8953.53	9217.30	9481.07
Step 50	4141.15	4224.75	4308.35	25359.72	25993.40	26627.08	8982.70	9243.10	9503.50





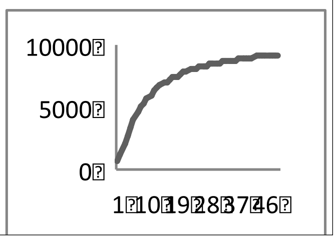


Fig. 63 Network total edge count development for ND = 6.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

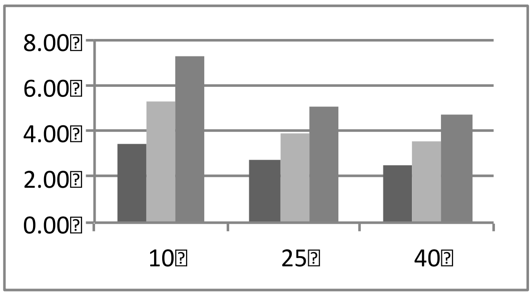
Step	AVG-STDEV	AVG	AVG+STDEV
10	3.44	5.38	7.31
25	2.75	3.94	5.14
40	2.47	3.60	4.73

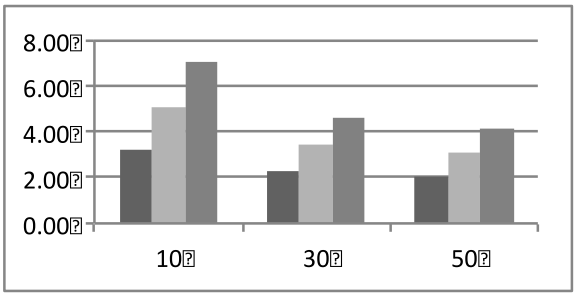
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.19	5.12	7.04
30	2.30	3.44	4.57
50	2.08	3.09	4.10

b.





a) Pathlength development for user(a) and article(b) networks when ND = 6

Average Clustering Coefficient in User Network

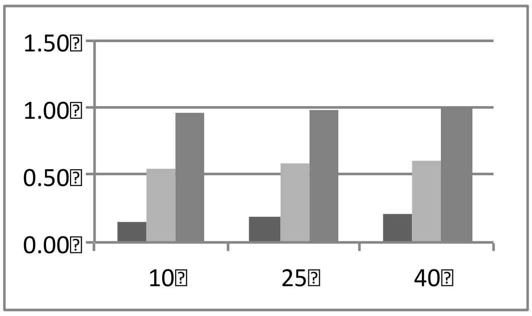
Step	AVG-STDEV	AVG	AVG+STDEV
10	0.15	0.55	0.96
25	0.20	0.59	0.99
40	0.21	0.60	1.00

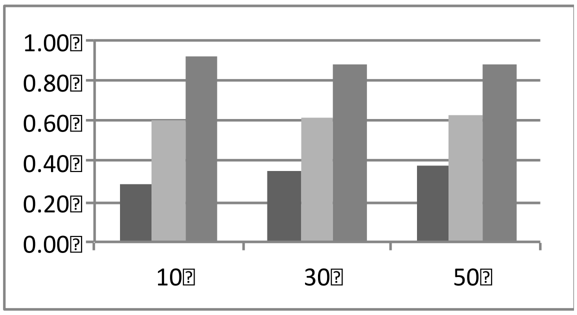
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.28	0.60	0.91
30	0.35	0.62	0.89
50	0.38	0.63	0.88

b.





b) Clustering coefficient development for user(a) and article(b) networks when ND = 6

Average Small World Coefficient in User Network

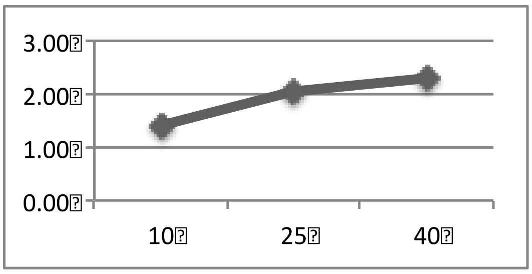
Step	AVG
10	1.41
25	2.07
40	2.31

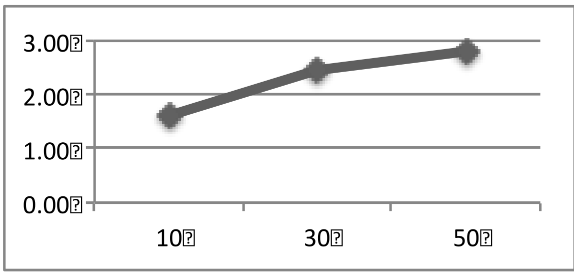
a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.61
30	2.48
50	2.80

b.

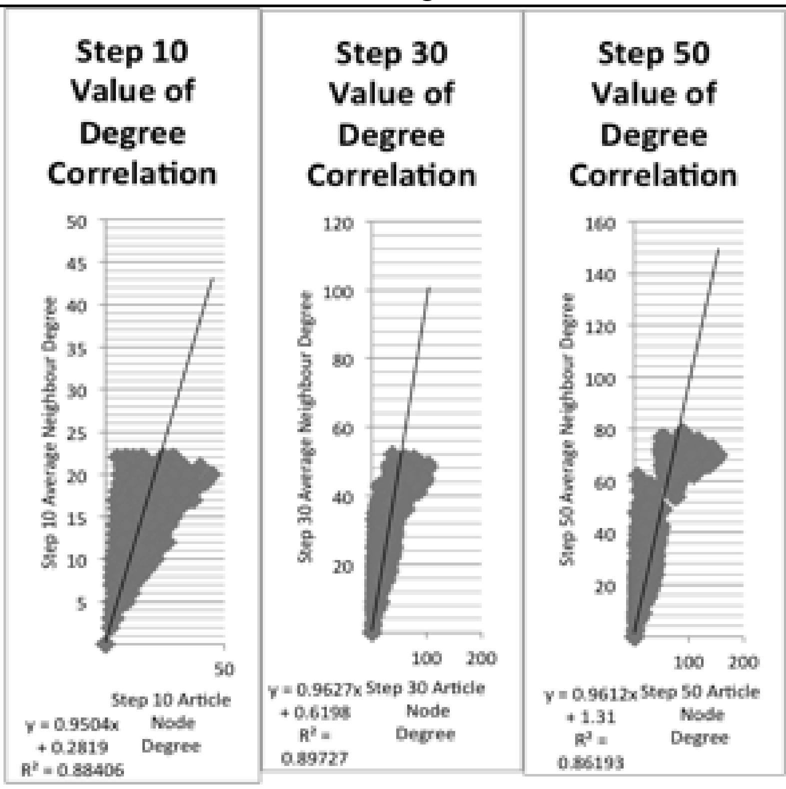




c) Small World coefficient development for user(a) and article(b) networks when ND = 6

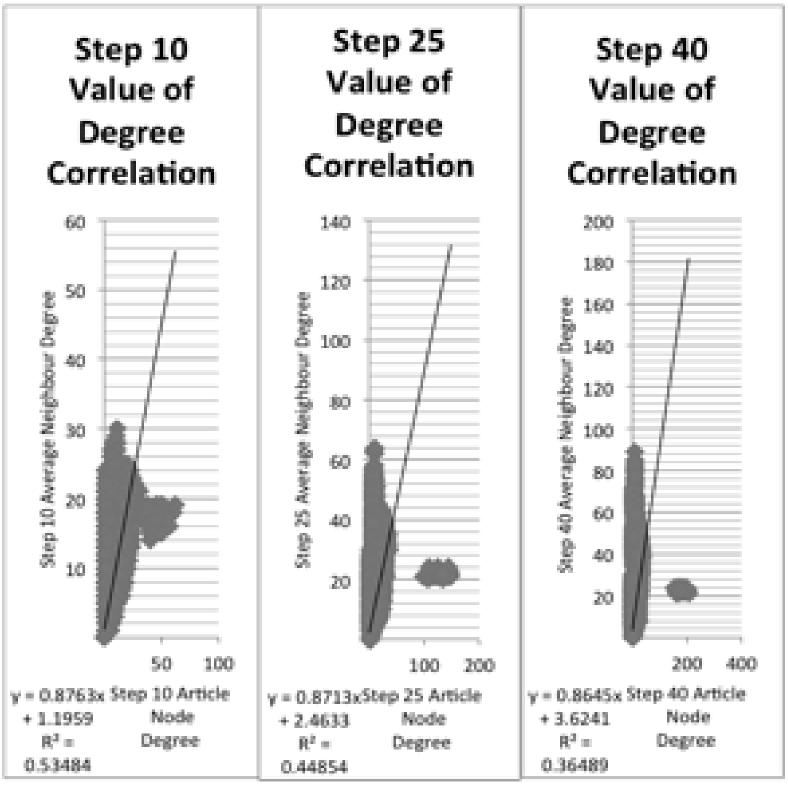
Fig. 64 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev. ND=6

Article Network Degree Correlation Values



a) Degree distribution development for article network when ND = 6

User Network Degree Correlation Values

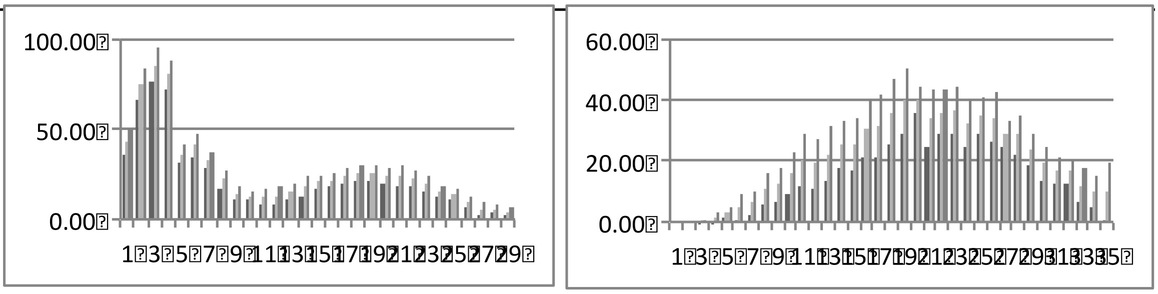


b) Degree distribution development for user network when ND = 6

Fig. 65 Degree correlation development for ND = 6.

Table 27 Article and User Network Degree Distribution for ND =.6

Node Degree	User Network User Count Having That Degree			Node Degree	Article Network Article Count Having That Degree		
	AVG-STDEV	AVG	AVG+STDEV		AVG-STDEV	AVG	AVG+STDEV
1	35.54	43.15	50.76	1	0.00	0.00	0.00
2	67.20	75.65	84.10	2	0.00	0.00	0.00
3	76.53	85.75	94.97	3	-0.21	0.17	0.54
4	72.44	80.75	89.06	4	-0.11	1.50	3.11
5	30.83	36.35	41.87	5	1.02	2.83	4.65
6	35.02	41.60	48.18	6	0.11	4.75	9.39
7	28.57	32.85	37.13	7	2.68	6.17	9.65
8	17.13	22.10	27.07	8	5.83	10.75	15.67
9	11.48	14.75	18.02	9	6.65	12.25	17.85
10	10.48	12.85	15.22	10	8.76	15.92	23.08
11	7.77	12.10	16.43	11	11.83	20.33	28.84
12	8.49	13.10	17.71	12	11.31	19.50	27.69
13	11.12	15.40	19.68	13	13.24	22.33	31.43
14	13.12	18.55	23.98	14	17.92	25.50	33.08
15	17.36	20.65	23.94	15	16.98	25.75	34.52
16	18.15	22.00	25.85	16	21.23	30.92	40.61
17	19.45	24.40	29.35	17	21.53	31.58	41.64
18	21.62	25.60	29.58	18	25.22	36.25	47.28
19	21.69	26.15	30.61	19	29.08	39.92	50.75
20	19.63	24.40	29.17	20	36.31	40.33	44.36
21	18.17	23.85	29.53	21	25.08	34.17	43.26
22	18.92	23.05	27.18	22	28.80	36.17	43.54
23	15.48	20.20	24.92	23	29.29	36.75	44.21
24	11.92	15.00	18.08	24	24.30	32.08	39.87
25	10.53	13.55	16.57	25	28.68	34.92	41.15
26	6.64	9.75	12.86	26	26.39	34.58	42.77
27	2.93	6.00	9.07	27	24.33	29.00	33.67
28	3.33	5.85	8.37	28	22.48	28.67	34.85
29	1.83	4.50	7.17	29	18.99	24.17	29.34
30	1.10	3.00	4.90	30	13.80	19.25	24.70
31	0.03	1.30	2.57	31	12.22	16.83	21.45
				32	12.35	16.50	20.65



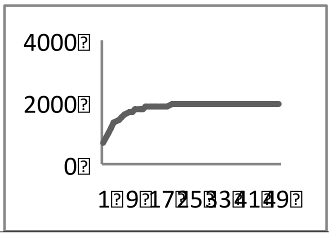
Degree distribution of Article(a) and User(b) networks when ND=.6.

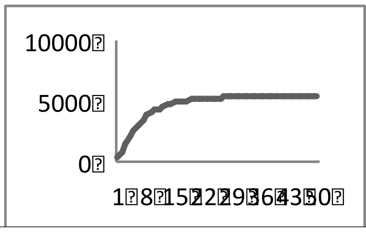
Fig. 66 Degree distribution for ND = 6.

Neighbourhood Dimension (ND) = 3, n=20 Replications Results

Table 28 Total Edge Count Development for ND = 3

	Affilitation Network			Article Network			User Network		
	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV	AVG - STDEV	AVG	AVG + STDEV
Step 1	673.87	686.25	698.63	281.60	289.60	297.60	637.50	667.25	697.00
Step 2	950.54	971.30	992.06	872.74	905.15	937.56	1064.50	1121.15	1177.80
Step 3	1147.58	1173.00	1198.42	1493.42	1549.20	1604.98	1415.10	1497.25	1579.40
Step 4	1292.56	1323.55	1354.54	2059.28	2144.00	2228.72	1698.79	1790.55	1882.31
Step 5	1405.40	1439.70	1474.00	2552.33	2655.50	2758.67	1916.17	2012.45	2108.73
Step 6	1496.77	1529.90	1563.03	2999.00	3086.25	3173.50	2086.45	2186.25	2286.05
Step 7	1565.94	1602.60	1639.26	3349.82	3455.35	3560.88	2205.41	2320.05	2434.69
Step 8	1624.60	1662.45	1700.30	3670.98	3768.50	3866.02	2312.91	2433.90	2554.89
Step 9	1672.72	1710.05	1747.38	3909.04	4011.65	4114.26	2406.72	2527.10	2647.48
Step 10	1711.15	1751.10	1791.05	4123.26	4236.50	4349.74	2478.01	2605.05	2732.09
Step 11	1740.44	1782.35	1824.26	4286.85	4411.05	4535.25	2525.03	2654.60	2784.17
Step 12	1763.01	1806.65	1850.29	4415.97	4552.30	4688.63	2559.25	2695.80	2832.35
Step 13	1780.16	1825.60	1871.04	4525.50	4669.00	4812.50	2593.12	2729.80	2866.48
Step 14	1795.92	1842.80	1889.68	4620.43	4770.90	4921.37	2618.65	2758.60	2898.55
Step 15	1809.02	1857.20	1905.38	4707.57	4860.95	5014.33	2636.04	2781.15	2926.26
Step 16	1821.52	1870.40	1919.28	4780.39	4943.30	5106.21	2657.44	2804.60	2951.76
Step 17	1828.24	1879.50	1930.76	4823.08	4995.65	5168.22	2668.46	2821.20	2973.94
Step 18	1837.25	1888.55	1939.85	4882.18	5052.70	5223.22	2685.50	2839.55	2993.60
Step 19	1845.38	1896.60	1947.82	4932.11	5102.95	5273.79	2697.53	2853.55	3009.57
Step 20	1851.36	1902.50	1953.64	4974.08	5140.55	5307.02	2709.47	2863.95	3018.43
Step 21	1856.20	1907.30	1958.40	5004.44	5172.35	5340.26	2716.33	2871.45	3026.57
Step 22	1860.05	1912.00	1963.95	5027.13	5204.90	5382.67	2722.48	2878.15	3033.82
Step 23	1863.30	1915.50	1967.70	5042.26	5223.90	5405.54	2726.33	2882.75	3039.17
Step 24	1866.80	1919.20	1971.60	5069.01	5251.90	5434.79	2730.93	2887.30	3043.67
Step 25	1869.49	1921.90	1974.31	5082.57	5269.25	5455.93	2733.69	2890.20	3046.71
Step 26	1871.10	1923.70	1976.30	5094.16	5282.00	5469.84	2736.46	2893.10	3049.74
Step 27	1872.47	1925.65	1978.83	5106.35	5294.75	5483.15	2738.13	2896.50	3054.87
Step 28	1874.28	1927.70	1981.12	5118.15	5307.80	5497.45	2740.35	2899.30	3058.25
Step 29	1875.95	1929.05	1982.15	5127.70	5316.45	5505.20	2743.19	2901.80	3060.41
Step 30	1877.48	1930.75	1984.02	5142.15	5329.70	5517.25	2745.68	2905.10	3064.52
Step 31	1878.77	1931.85	1984.93	5150.64	5336.40	5522.16	2747.91	2906.80	3065.69
Step 32	1880.10	1933.15	1986.20	5158.22	5345.55	5532.88	2750.15	2908.65	3067.15
Step 33	1880.38	1933.60	1986.82	5160.52	5347.95	5535.38	2750.54	2909.05	3067.56
Step 34	1881.13	1934.40	1987.67	5167.44	5354.60	5541.76	2751.09	2910.25	3069.41
Step 35	1881.67	1935.25	1988.83	5170.40	5361.25	5552.10	2751.16	2910.95	3070.74
Step 36	1882.35	1935.70	1989.05	5176.40	5364.95	5553.50	2751.28	2911.15	3071.02
Step 37	1882.59	1936.10	1989.61	5179.09	5367.15	5555.21	2751.65	2911.55	3071.45
Step 38	1883.35	1936.75	1990.15	5183.12	5371.45	5559.78	2753.00	2912.75	3072.50
Step 39	1883.90	1937.10	1990.30	5185.06	5373.10	5561.14	2753.02	2913.45	3073.88
Step 40	1883.90	1937.10	1990.30	5185.06	5373.10	5561.14	2753.02	2913.45	3073.88
Step 41	1884.28	1937.55	1990.82	5188.77	5376.15	5563.53	2753.38	2913.85	3074.32
Step 42	1884.71	1937.80	1990.89	5192.35	5378.20	5564.05	2753.53	2913.95	3074.37
Step 43	1884.86	1937.95	1991.04	5194.38	5379.65	5564.92	2753.72	2914.15	3074.58
Step 44	1885.30	1938.30	1991.30	5200.30	5383.25	5566.20	2753.54	2914.75	3075.96
Step 45	1885.30	1938.35	1991.40	5200.59	5383.55	5566.51	2753.51	2914.85	3076.19
Step 46	1885.34	1938.45	1991.56	5200.71	5383.85	5566.99	2753.68	2914.95	3076.22
Step 47	1885.63	1938.80	1991.97	5202.46	5386.70	5570.94	2754.15	2915.35	3076.55
Step 48	1885.73	1938.90	1992.07	5203.54	5387.70	5571.86	2754.15	2915.35	3076.55
Step 49	1885.77	1939.00	1992.23	5203.99	5389.40	5574.81	2754.30	2915.55	3076.80
Step 50	1885.89	1939.20	1992.51	5205.87	5391.95	5578.03	2754.49	2915.75	3077.01





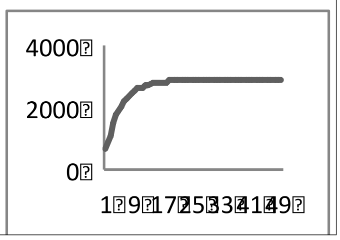


Fig. 67 Network total edge count development for ND = 3.

Key Network Metrics for Article and User Networks

Average Pathlength in User Network

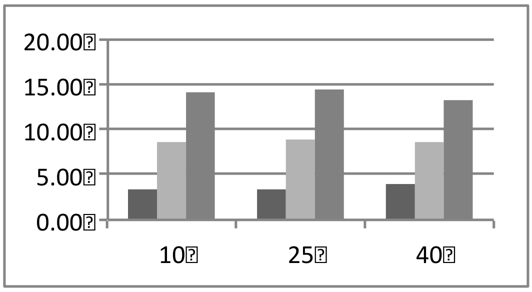
Step	AVG-STDEV	AVG	AVG+STDEV
10	3.44	8.73	14.03
25	3.41	8.97	14.53
40	3.98	8.62	13.27

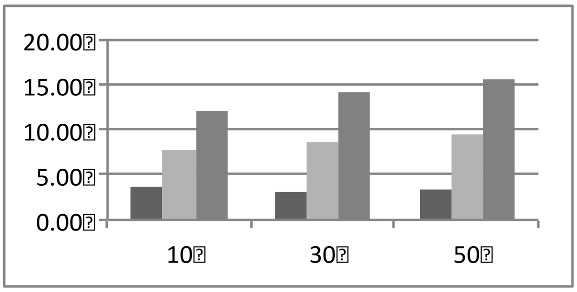
a.

Average Pathlength in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	3.58	7.84	12.10
30	3.07	8.58	14.09
50	3.43	9.48	15.53

b.





a) Pathlength development for user(a) and article(b) networks when ND = 3.

Average Clustering Coefficient in User Network

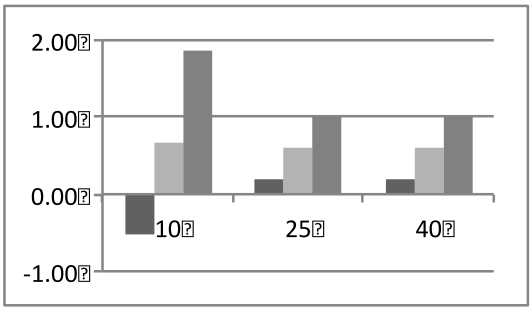
Step	AVG-STDEV	AVG	AVG+STDEV
10	-0.54	0.67	1.87
25	0.18	0.60	1.03
40	0.18	0.60	1.03

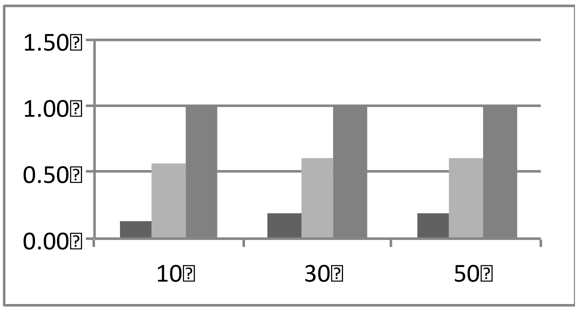
a.

Average Clustering Coefficient in Article Network

Step	AVG-STDEV	AVG	AVG+STDEV
10	0.14	0.57	1.00
30	0.19	0.60	1.01
50	0.19	0.60	1.01

b.





b) Clustering coefficient development for user(a) and article(b) networks when ND = 3.

Average Small World Coefficient in User Network

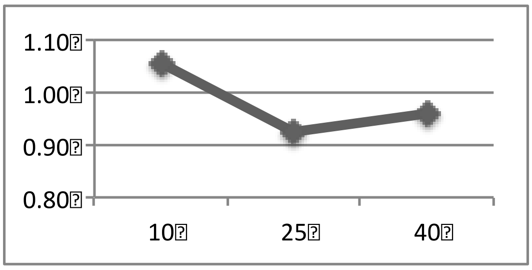
Step	AVG
10	1.05
25	0.92
40	0.96

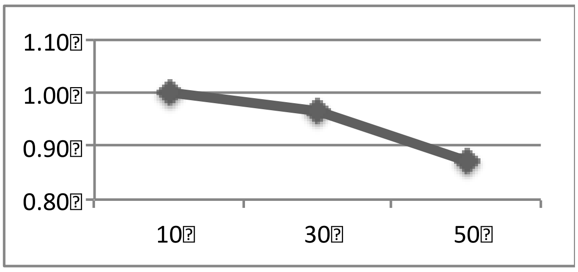
a.

Average Small World Coefficient in Article Network

Step	AVG
10	1.00
30	0.96
50	0.87

b.

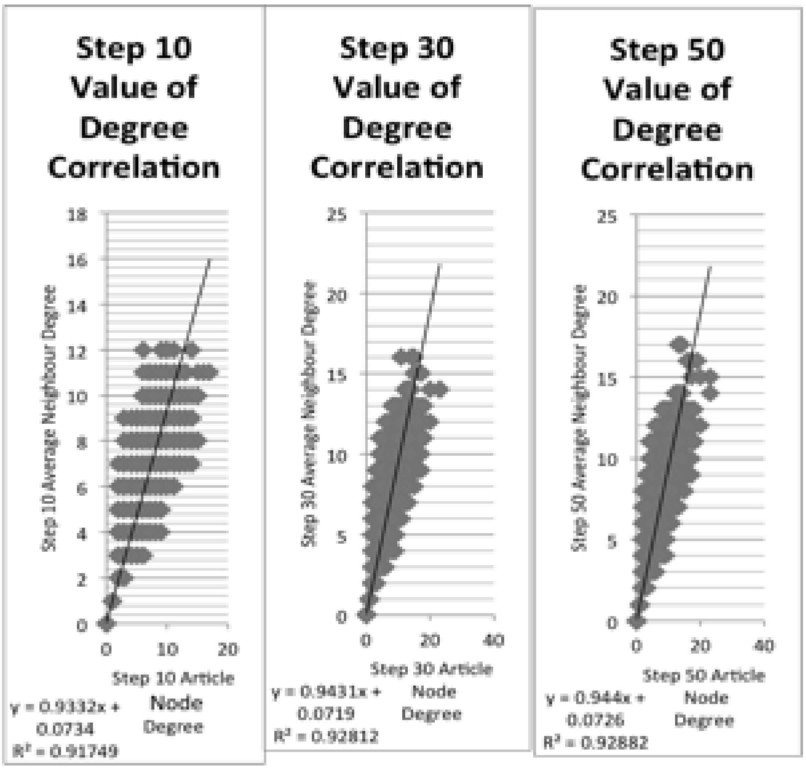




c) Small World coefficient development for user(a) and article(b) networks when ND = 3.

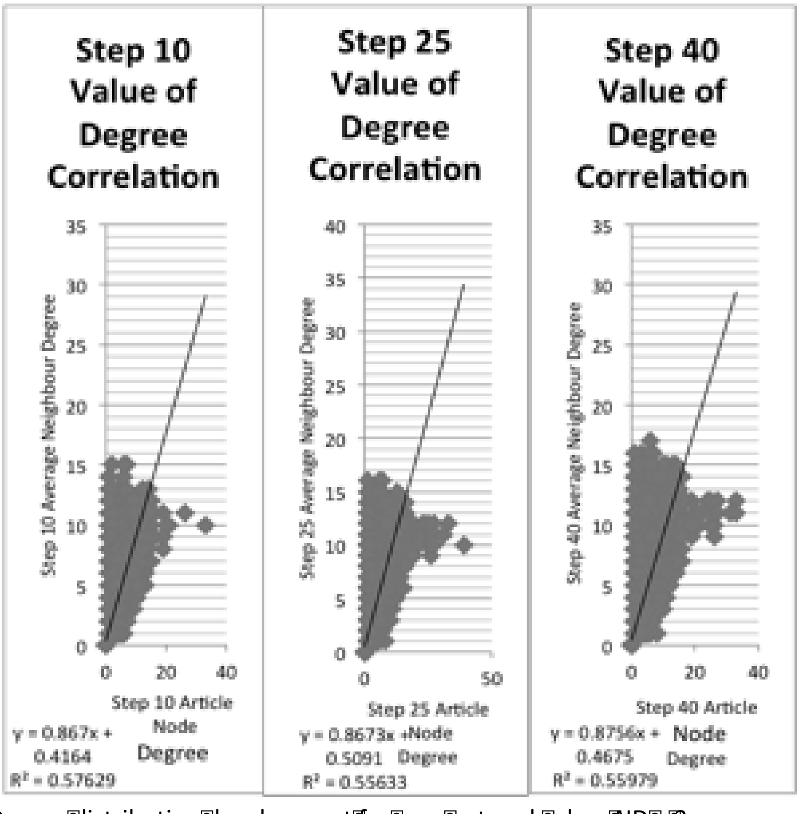
Fig. 68 Path length (a), Clust. Coeff. (b), Small-World Char. Q (c) dev ND=3.

Article Network Degree Correlation Values



a) Degree distribution development for article network when ND = 3.

User Network Degree Correlation Values



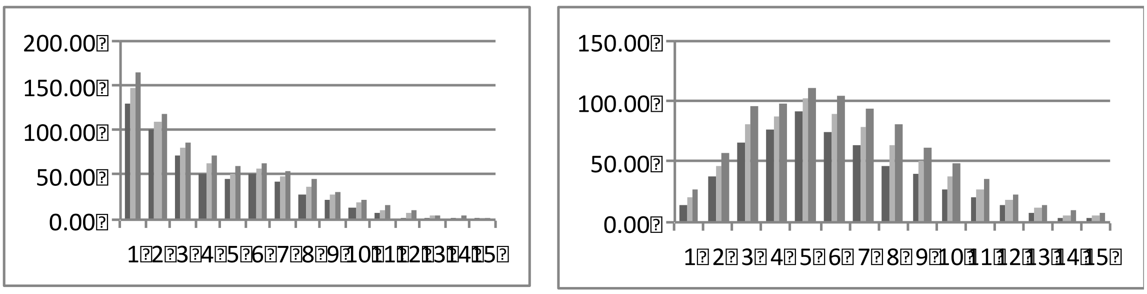
b) Degree distribution development for user network when ND = 3.

Fig. 69 Degree correlation development for ND = 3.

Table 29 Article and User Network Degree Distribution for ND = 3

Node Degree	User Network User Count Having That Degree AVG-STDEV	AVG	AVG= STDEV	Node Degree	Article Network Article Count Having That Degree AVG-STDEV	AVG	AVG= STDEV
1	128.82	146.50	164.18	1	14.38	20.83	27.29
2	101.68	109.80	117.92	2	37.18	46.83	56.48
3	72.40	80.00	87.60	3	65.46	80.83	96.21
4	52.32	62.00	71.68	4	77.41	88.25	99.09
5	44.98	52.45	59.92	5	92.80	102.50	112.20
6	52.10	57.60	63.10	6	74.88	90.25	105.62
7	41.52	47.75	53.98	7	63.64	78.92	94.19
8	29.17	37.50	45.83	8	47.56	64.75	81.94
9	21.92	27.10	32.28	9	40.81	51.67	62.52
10	13.56	18.30	23.04	10	27.52	38.00	48.48
11	6.70	11.05	15.40	11	19.93	27.50	35.07
12	3.35	6.60	9.85	12	14.51	18.58	22.65
13	2.14	3.85	5.56	13	8.71	11.67	14.62
14	0.27	2.60	4.93	14	2.91	6.33	9.76
15	0.37	1.65	2.93	15	2.87	5.25	7.63

a. b.



Degree distribution of User(a) and Article(b) networks.

Fig. 70 Degree distribution for ND = 3.

BIBLIOGRAPHY

Abe. (2010). Diving into the deep end of ABMS. In agentbasedmodeling.wordpress.com. Retrieved March 20, 2013 from http://agentbasedmodeling.wordpress.com. Also see: http://3.bp.blogspot.com/-8pFTFfbEJsE/TxSj3FfwciI/AAAAAAAAAQA/Lp8JvncRKXA/s1600/dont-use-netlogo.png

Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the Internet. *Glottometrics*, *3*(1), 143-150.

Alexa. (2013). Wikipedia Traffic Stats [Graph illustration]. Retrieved from http://www.alexa.com/siteinfo/en.Wikipedia.org

Anthony, D., Smith, S. W., & Williamson, T. (2005). Explaining quality in internet collective goods: Zealots and good samaritans in the case of Wikipedia. *Hanover: Dartmouth College*.

Anthony, D., Smith, S. W., & Williamson, T. (2007). The quality of open source production: Zealots and good samaritans in the case of Wikipedia. *Rationality and Society*.

Aumann, R. J. (1959). Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games*, *4*, 287-324.

Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of conflict resolution*, *24*(1), 3-25.

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton: Princeton University Press.

Axelrod, R. (1997b). Advancing the Art of Simulation in the Social Sciences, *Complexity* 3. 193–199.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2012, June). Four degrees of separation. In *Proceedings of the 3rd Annual ACM Web Science Conference* (pp. 33-42). ACM.

Barabási, A. L. & Albert R. (1999). Emergence of scaling in random networks. *Science 286* (5439): 509–512.

Blumenstock, J. E. (2008, April). Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web* (pp. 1095-1096). ACM.
Broad, C.D. (1925). *The Mind and its Place in Nature*. London: Routledge and

Kegan Paul, Ltd.

Casebourne, I., Davies, C., Fernandes, M., Norman, N. (2012). Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic. Epic, Brighton, UK. Retrieved from: http://commons.wikimedia.org/wiki/File:EPIC_Oxford_report.pdf

Cauchy, A. L. (1813). Sur les polygones et polyedres, Second mémoire. *J. Ecole Polytechnique,* 9, 87.

Cioppa, T. M., Lucas, T. W., & Sanchez, S. M. (2004, December). Military applications of agent-based simulations. In *Simulation Conference, 2004. Proceedings of the 2004 Winter (Vol. 1)*. IEEE.

Collier, N., Howe, T., & North, M. (2003, June). Onward and upward: The transition to Repast 2.0. In *First Annual North American Association for Computational Social and Organisational Science Conference,* Pittsburgh, PA.

Conway, J. (1970). The game of life. *Scientific American, 223*(4), 4.

Diderot, D., & d'Alembert, J. L. R. (1751). *Encyclopédie, ou Dictionnaire raisonné des Sciences, des Arts et des Métiers.* (Vol. 2). Paris: chez Briasson.

Delre, S. A., Jager, W., Bijmolt, T. H. A., & Janssen, M. A. (2007). Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of business research*, *60*(8), 826-835.

Emrich, S., Breitenecker, F., Zauner, G., & Popper, N. (2008, June). Simulation of influenza epidemics with a hybrid model-combining cellular automata and agent based features. In *Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on* (pp. 709-714). IEEE.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae 8,* 128-140.

Flood, A. (2012). Encyclopedia Britannica's final print edition on verge of selling out. Retrieved September 14, 2014, from http://www.theguardian.com/books/2012/apr/05/encyclopedia-britannica-final-print-edition

Getchell, A. (2008). Agent-based modeling. *Physics,* 22(6), 757-767.

Giardina, I., & Bouchaud, J. (2003). Bubbles, crashes and intermittency in agent based market models. *The European Physical Journal B-Condensed Matter and Complex Systems, 31*(3), 421-437.

Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind Society, 1*(1), 57-72. Springer.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature. 438(7070), 900-901 (December 2005).*

Google Trends. (2013). Web Search Interest: agent based modelling, agent-based model. Worldwide, 2004-present. [Graph illustration]. Retrieved March, 17, 2013, from http://www.google.com/trends/explore#q=agent%20based%20modelling%2C%20agent-based%20model&cmpt=q

Gray, A. (2010, November). What proportion of articles are stubs? Retrieved September 13, 2014, from http://wikimedia.7.x6.nabble.com/What-proportion-of-articles-are-stubs-td1107601.html

Grimm, V., & Railsback, S. F. (2013). *Individual-based modeling and ecology*. Princeton: Princeton University Press.

Hernández Encinas, L., Hoya White, S., Martín del Rey, A., & Rodríguez Sánchez, G. (2007). Modelling forest fire spread using hexagonal cellular automata. *Applied mathematical modelling, 31*(6), 1213-1227.

Holland, J. H. (1995). *Hidden Order: How Adaptation Builds Complexity*. United States of America: Perseus Books.

Holmgren, J., Persson, J. A., & Davidsson, P. (2013). Improving multi-actor production, inventory and transportation planning through agent-based optimization. In *Agent and Multi-Agent Systems in Distributed Systems-Digital Economy and E-Commerce* (pp. 1-31). Springer Berlin Heidelberg.

Ingawale, M. (2008, October). Understanding the Wikipedia phenomenon: a case for agent based modeling. In *Proceedings of the 2nd PhD workshop on Information and knowledge management* (pp. 81-84). ACM.

Ingawale, M., Dutta, A., Roy, R., & Seetharaman, P. (2009, January). The small worlds of Wikipedia: Implications for growth, quality and sustainability of collaborative knowledge networks. In *Proceedings of the 15th Americas conference on information systems.* Retrieved from http://aisel. aisnet. org/amcis2009/439.

Izquierdo, L. R., Galán, J. M., Santos, J. I., del Olmo, R. (2008). Modelado de sistemas complejos mediante simulación basada en agentes y mediante dinámica de sistemas. *Empiria: Revista de Metodología de Ciencias Sociales*, (16), 85-112.

Kamalabadi, E. (2006). Wikipedia:Wikipediology/library/essays/Exir Kamalabadi-1 ... Retrieved September 11, 2014, from http://webcache.googleusercontent.com/search?q=cache:J7PiwdhgoFcJ:www.quickiwiki.com/en/Wikipedia:Wikipediology/library/essays/Exir_Kamalabadi-1+&cd=1&hl=tr&ct=clnk&gl=tr&client=firefox-a

Kesting, A., Treiber, M., & Helbing, D. (2008). Agents for traffic simulation. *arXiv preprint arXiv:0805.0300.*

Lattanzi, S., & Sivakumar, D. (2009, May). Affiliation networks. In *Proceedings of the 41st annual ACM symposium on Theory of computing* (pp. 427-434). ACM.

Lewin, R. (1992). *Complexity: Life at the Edge of Chaos*: Macmillan.

Lohmann, E., Köroğlu, Ç., Hanagasi, H. A., Dursun, B., Taşan, E., & Tolun, A. (2012). A homozygous frameshift mutation of sepiapterin reductase gene causing parkinsonism with onset in childhood. *Parkinsonism & related disorders*, *18*(2), 191-193.

Lytinen, S. L., & Railsback, S. F. (2012). The Evolution of Agent-based Simulation Platforms: A Review of NetLogo 5.0 and ReLogo. In *Proceedings of the Fourth International Symposium on Agent-Based Modeling and Simulation.*

Macal, C. M., & North, M. J. (2008, December). Agent-based modeling and simulation: ABMS examples. In *Proceedings of the 40th Conference on Winter Simulation* (pp. 101-112). Winter Simulation Conference.

McMullin, B. (2000). John von Neumann and the evolutionary growth of complexity: Looking backward, looking forward. *Artificial Life*, *6*(4), 347-361.

Milgram, S. (1967). The small-world problem. *Psychology Today*, 61-67.

Miller, J. H. (1986). A genetic model of adaptive economic behaviour. Working Paper. *Ann Arbor:* University of Michigan.

Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008, January). On the inequality of contributions to Wikipedia. In proceedings of 41st Annual Hawaii International Conference on System Sciences. Piscataway, NJ: IEEE Press. Retrieved October 5, 2014 from http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.333

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation, 82*(9), 609-623.

Reavley, N. J., Mackinnon, A. J., Morgan A. J., Alvarez-Jimenez M., Hetrick S.E., Killackey E., Nelson B., Purcell R., Yap M.B., Jorm A.F. (2012). "Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources." *Psychological medicine* 42.8 (2012): 1753.

Schelling, T. C. (1971). Dynamic models of segregation†. *Journal of mathematical sociology, 1*(2), 143-186.

Shuangyun, P., Kun, Y., Quanli, X., & Yuhua, H. (2010). The Simulation Research on the Evacuation Model of Supermarket by Using GIS and Agent-Based Modeling Technology. *Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium on*. IEEE.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70*(5), 893-912.

Smith, E. R. (1998). Mental representation and memory. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 391-445). New York: McGraw-Hill.

Solomon, S. (2013). Common Complex Collective Phenomena: implications for economic and social policy making. Retrieved April 9, 2013, from http://www.sorinsolomon.net/~sorin/ccs/BroszuraCO3B(2).doc.

Statistics. (n.d.). In Wikipedia. Retrieved July 15, 2014, from http://en.wikipedia.org/wiki/Special:Statistics

Tesfatsion, L., & Axelrod, R. On-Line Guide for Newcomers to Agent-Based Modeling in the Social Sciences. Retrieved April 10, 2013, from http://www2.econ.iastate.edu/tesfatsi/abmread.htm

Turner, A., & Penn, A. (2002). Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environ Plann B, 29*(4), 473-490.

User: Opabinia regalis/Article_statistics. (2007, 18 February). In *Wikipedia*. Retrieved September 9, 2014, from http://en.wikipedia.org/w/index.php?title=User:Opabinia_regalis/Article_statistics&oldid=294953437

Viégas, F. B., Wattenberg, M., & Dave, K. (2004, April). Studying cooperation and conflict between authors with history flow visualisations. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 575-582). ACM.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440-442.

Waugh, R. (2012, April 20). First man hits a MILLION Wikipedia edits after averaging an incredible 385 per day for seven years. *Daily Mail*. Retrieved from http://www.dailymail.co.uk

What is WAM Score?. (2014). Retrieved September 11, 2014, from http://www.wikia.com/WAM

Wilkinson, D. M., & Huberman, B. A. (2007, October). Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis* (pp. 157-164). ACM.

Wikipedia: Statistics. (n.d.). In *Wikipedia*. Retrieved May 4, 2013, from http://en.wikipedia.org/w/index.php?title=Wikipedia:Statistics&oldid=553329592

Wikipedia: About. (n.d.). In *Wikipedia*. Retrieved September 9, 2014, from http://en.wikipedia.org/w/index.php?title=Wikipedia:About&oldid=617443622

Wikipedia: Why create an account?. (n.d.). In *Wikipedia*. Retrieved September 9, 2014, from http://en.wikipedia.org/w/index.php?title= Wikipedia:Why_create_an_account%3F&oldid=624539794

Wikipedia:Good Articles. (2014, September). In *Wikipedia*. Retrieved September 10, 2014, from http://en.wikipedia.org/w/index.php?title=Wikipedia:Good_articles&oldid=608372225

Wöhner, T., & Peters, R. (2009, October). Assessing the quality of Wikipedia articles with lifecycle based metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (p. 16). ACM.

Yıldırım, M. (2009). Building an agent based model for observing development of a web 2.0 project (Master's thesis). Retrieved from http://knuth.cs.bilgi.edu.tr/~myildirim/mla/cs_thesis.pdf

Zachte, E. (2014, September 12). Wikipedia Statistics All languages. Retrieved September 12, 2014, from http://stats.wikimedia.org/EN/TablesWikipediaZZ.htm