

Identifying High-status Nodes in Knowledge Networks

Siddharth Kaza and Hsinchun Chen

Department of Computer and Information Sciences, Towson University, Towson, MD, USA, Department of Management Information Systems, University of Arizona, Tucson, AZ, USA

skaza@towson.edu, hchen@eller.arizona.edu

Abstract The status of a node in a social network plays an important part in determining evolution of the network around it. High-status nodes in knowledge networks are likely to attract more links and influence the use of knowledge by nodes connected directly or indirectly to them. In this study, we model knowledge flow within an innovative organization and contend that it exhibits unique characteristics not incorporated in most social network measures designed to determine node status. Based on the model, we propose the use of a new measure based on team identification and random walks to determine status in knowledge networks. Using data obtained on collaborative patent networks, we find that the new measure performs better than others in identifying high-status inventors.

1. Introduction

The status of a node in a social network can determine its evolution and influence other nodes linked directly or indirectly to it. In knowledge networks – defined here as social networks with individuals sharing knowledge with each other while being connected through collaborative links – the status of individuals can influence the evolution of knowledge and the innovation resulting from it. Innovation has been described as a problem solving process where the solutions are discovered via the search novel recombination of existing knowledge (Fleming, 2001; Kogut & Zander, 1992). Organizations have a choice in selecting knowledge that is recombined to produce new innovations. The selection of knowledge for recombination is influenced by the status of inventors in an organization’s internal knowledge network (Nerkar & Paruchuri, 2005; Singh, 2007; Tsai, 2001). Organizations (and inventors within) attach more value and recombine knowledge of high-status inventors.

Various social network measures have been used to establish the status of inventors in knowledge networks (Newman, 2004; Patrakosol & Olson, 2007; Singh, 2005; Sorenson & Stuart, 2001). However, the measures make implicit assumptions about the flow of knowledge within an organization. For instance, the widely used betweenness centrality measure (Freeman, 1979) assumes that knowledge flows along shortest paths. Often these assumptions are not valid for modeling knowledge flow and establishing the status of inventors based on these measures may lead to misleading results.

In this paper, we determine the inventor status in intra-organizational knowledge network and study its effect on the selection of knowledge that is recombined to produce innovation. We focus on intra-organization networks since recombination of internal knowledge helps establish competitive advantage for a longer time (Chesbrough & Tece, 1996; Rosenkopf & Nerkar, 2001). We model the flow of knowledge within a research focused organization and contend that it exhibits unique characteristics not incorporated in most social network measures. Using the model, we also propose a new measure based on random walks and team identification and use it to examine innovation selection in a large organization.

In particular, we explore the following research questions: How can we effectively model the flow of knowledge within an intra-organizational knowledge network? How can we establish the status of an individual in a collaborative knowledge network? How does the status of an inventor in a knowledge network affect innovation evolution?

The rest of this paper is organized as follows: Section 2 presents the literature review and background, Section 3 describes the research design and testbed. Section 4 presents the experimental results and discussion. Section 5 concludes and proposes future directions.

2. Literature Review

2.1 Social network measures

Various measures to quantify characteristics of social networks have been proposed in the literature (Albert & Barabasi, 2002; Wasserman & Faust, 1994). Small world measures (clustering coefficient and average shortest path length) are used to examine the structure of the whole network. The clustering coefficient shows the tendency of individuals to cluster together to form cliques. The average shortest path length shows that even though a network may be large, most individuals are located within a few steps from each other. However, in this section, we focus on measures that are used to identify high-status nodes in a network (these measures are usually known as centrality or prestige measures). As mentioned before, these high-status individuals may influence the evolution of innovation in an organization.

Several measures of node centrality have been developed including degree centrality, closeness, betweenness, information centrality, and influence measures (Borgatti, 2005; Scott, 2000). These measures are not independent of the dynamic processes that unfold within a network (Friedkin, 1991) and make different implicit assumptions about the path of knowledge flow in a network. For instance, Freeman's betweenness centrality assumes that knowledge flows along shortest paths in the network (Freeman, 1979). However, many studies use these measures without regard to the implicit assumptions made by them. This might lead to poor results or a wrong interpretation of the network phenomenon under study (Borgatti, 2005). Thus, it is necessary to model the assumptions pertinent to the network under study prior to selecting the centrality measure.

Based on analysis of previous studies (Borgatti, 2005; Fleming, King, & Juda, 2007; Uzzi & Spiro, 2005), we contend that there are three primary requirements for a measure to correctly identify high-status nodes in a knowledge network of inventor collaborations. These are:

Account for Diversity of knowledge (D): This implies that a high status inventor is likely to receive diverse knowledge from different parts of the network. In SNA theory, this is best represented by betweenness measures. Betweenness is a measure of the influence a node has on the spread of information through the network (Newman, 2005). The higher the betweenness, the more frequently a node is likely to receive information from disjoint parts of a network. This is important as the recombination of diverse knowledge from disjoint parts of the network is likely to lead to more innovation (Fleming et al., 2007; Uzzi & Spiro, 2005).

Random diffusion (R): This implies that the measure should assume that knowledge does not select a preferred path (like the shortest path) of travel through a network. This does not necessarily imply that all paths (of all lengths) are equally important. It has been shown that shorter paths may be important in transferring certain kinds of knowledge (Hansen, 1999).

Parallel duplication (P): This implies that multiple copies of the same knowledge can exist in a network. Thus, when given a choice in the path of travel, knowledge can travel on multiple paths at once (Borgatti, 2005). For instance, knowledge is transferred to multiple individuals during team presentations. This assumption is especially important in this study since we are studying inventors within organizations where they are likely to be organized in project teams.

Table 1 reviews important centrality measures and classifies them according to these requirements. As shown in the table, none of the measures satisfy all the requirements to model information flow. For instance, Freeman's betweenness measure (Freeman, 1979) does not take into account the duplication of knowledge. Bonacich's power (Bonacich, 1987) accounts for random diffusion and parallel duplication, however, it is not a betweenness measure and thus does not consider diversity of knowledge. Newman's random walk betweenness (2005) assumes *D* and *R* however, does not contain a parallel duplication component. A comprehensive discussion of these centrality measures and their assumptions is provided by Borgatti (2005). We propose a measure based on Newman's random walk betweenness centrality to model knowledge flow in the collaboration networks studied here. A team identification component is added to the measure that assumes parallel duplication of knowledge within teams in an organization. Details of the proposed measure are presented in the research design. We believe that the proposed measure satisfies all three requirements for knowledge flow and better identifies high status inventors.

Table 1. Centrality measures and their knowledge flow assumptions

Measure	Intuition / Formulation	Knowledge Flow Assumption	Requirement
Closeness centrality (Freeman, 1979)	Sum of geodesic distances from all other nodes	Shortest path, no parallel duplication	
Betweenness centrality (Freeman, 1979)	Number of times that a node is on the shortest path between two nodes	Shortest path, no parallel duplication	D
Degree centrality (Freeman, 1979)	Node is central if it has a high degree	1-link paths, parallel duplication	P
Bonacich power (Bonacich, 1987)	Node is central to the extent that it is connected to other central nodes	Random diffusion, parallel duplication	R P
Information centrality (Stephenson & Zelen, 1989)	Harmonic mean of lengths of paths ending at a node	Random diffusion, no parallel duplication	R
Flow betweenness (Freeman, Borgatti, & White, 1991)	Flow through a node when a maximum amount of flow travels between source and target	Defined path, parallel duplication	D P
Structural holes (Burt, 1992)	Non-redundancy of a nodes neighbors and the links between them (usually 1-link neighborhood)	1-link paths, parallel duplication	D P
Random-walk centrality (Noh & Rieger, 2002)	Speed at which a random walk reaches a target node	Random diffusion, no parallel duplication	R
Random-walk betweenness centrality (Newman, 2005)	Number of times a random-walk between a source and target passes through a node	Random diffusion, no parallel duplication	D R

2.2 Innovation and knowledge networks

Innovation has been described in the literature as a problem-solving process wherein solutions are discovered via the search and recombination of existing knowledge (Fleming, 2001; Henderson & Cockburn, 1994; Kogut & Zander, 1992). During this process, each organization is faced with a decision to select existing knowledge that is recombined to produce new innovative artifacts. As the recombination process proceeds, a focal innovation emerges that other innovations build upon (Nerkar & Paruchuri, 2005; Podolny & Stuart, 1995; Rosenkopf & Nerkar, 2001). In order to understand innovation evolution, it is necessary to identify the factors leading to the selection of a focal innovation. It has been shown that individuals and organizations do not select innovations just by their technical merits (Fleming, III, & Juda, forthcoming; Podolny & Stuart, 1995), other factors like the expertise of inventors, scope of the innovation, and number of other innovations in the same field play an important role in the selection process. Inventors also select the focal innovation based on the status of the innovation's inventors in the knowledge network (Nerkar & Paruchuri, 2005; Singh, 2007). One way to establish the status of an inventor is to use social network measures.

Various studies (Table 2) have focused on the status of inventors and innovation selection. Singh (2007) found that the degree centrality of an inventor did not have a significant effect on the impact of his/her innovation. Podolny & Stuart (1995) found that inventor-status did not have a significant positive impact of the selection of their innovation. They also found that the status of other related innovators in the network had a positive association on the impact of the focal innovation. However, both these studies used degree to establish status which may not be an accurate representation of inventor status in a knowledge network and thus may not give the right results. Singh (2005) found that as shortest path length between inventors increased, they were less likely to cite each other. The study acknowledged that presence of multiple paths between inventors may have different effects. Nerkar & Paruchuri (2005) found that Bonacich power of an inventor had a significant positive impact on the selection of his/her innovation. We used the same statistical technique as their study, however, we proposed a new measure better suited to the problem domain.

Table 2. Studies using node-level measures

	Network Ex- tent	Measures	Aim/Result
(Breschi & Lissoni, forthcoming)	Inter-org.	Shortest path	Study the geographic diffusion of innovation
(Patrakosol & Olson, 2007)	Inter-org.	Degree centrality	Effect of collaboration on IT innovation. Result: Close collaborations lead to evolutionary innovation
(Singh, 2007)	Inter-org.	Degree centrality and extensions	Impact of collaboration on innovation selection and future productivity
(Bell, 2005)	Both	Degree centrality	Impact of managerial network on innovation. Result: higher degree leads to higher innovation
(Nerkar & Paruchuri, 2005)	Intra-org.	Bonacich power, structural holes	Impact of inventor positions on innovation selection.
(Singh, 2005)	Both	Shortest paths	Effect of shortest path on innovation selection
(Ahuja, 2000)	Inter-org.	Node degree, structural holes	Effect of measures on the organizations' innovative output. Result: degree – positive, structural holes – negative.
(Podolny & Stuart, 1995)	Inter-org.	Degree centrality	Study the factors that determine innovation selection

Studies marked in **bold** specifically focused on innovation selection. These are discussed in detail in the text.

3. Research Design and Testbed

Fig. 1 shows the research design and process used to acquire data, extract knowledge networks, develop the network measures, and statistically evaluate the effect of network measures on innovation selection.

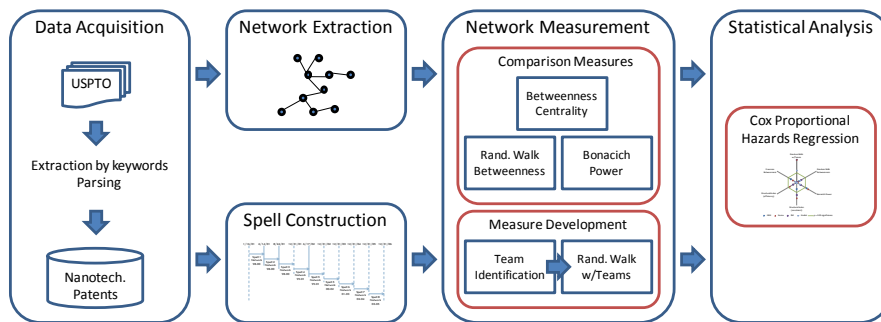


Fig. 1. Research design and process

3.1 Data acquisition

This study used nanotechnology related patents from the United States Patent and Trademark Office (USPTO). This is because patents are considered to be excellent indicators of innovation in organizations (Jaffe & Trajtenberg, 2002; Nerkar & Paruchuri, 2005). We selected nanotechnology as it is an innovative field that promises fundamental changes to a wide variety of research domains (Chen, Roco, Li, & Lin, 2008). The patents were limited to the nanotechnology field by using a keyword search on the full text of the patent (for details see Huang, Chen, Chen, & Roco (2004)). Each patent document was downloaded using a web spider and parsed to extract information on assignee organization, inventors, issue and application dates, citation, and other fields. Table 3

shows the statistics of the patents obtained. The testbed in this study included the top organization by the number of inventors (International Business Machines – IBM). Large organizations are usually in business for a longer period of time and tend to have more established knowledge networks and better developed internal knowledge. This is important in this study as an organization with a quality internal knowledge base is likely to specialize in a certain area and recombine its own knowledge to produce innovations.

Table 3. Key statistics of nanotechnology patents extracted from USPTO

Date range	1976-2006
Patents	97, 562
Assignee institutions (organizations)	26, 304
Inventors (individuals)	189, 045

3.2 Network extraction

A knowledge network based on common affiliations was extracted for inventors in IBM. In the network, each node was represented by an inventor and two inventors were linked to each other if they were listed on the same patent. Such a network reflects strong associations as inventors listed on the same patent are likely to have intense collaboration while working on that innovation. Such an observed collaboration marks the beginning of a strong tie that lasts beyond the collaboration date (Agrawal, Cockburn, & McHale, 2006; Singh, 2005).

3.3 Spell construction

A spell divides the life of a patent (from issue date till the end of the dataset) into time periods. Each time period is used as a data point to determine the effect of various variables on the citation (or no citation) of the patent in that spell. In line with prior research (Nerkar & Paruchuri, 2005; Podolny & Stuart, 1995), spells of up to 1 year were created for each patent. The first spell began at issue date and ended at either the close of the same year or at the citation date if the patent is cited within that year. The next spell began at the start of the year - if the previous spell ended at the previous year or at the citation date - if the previous spell ended in a citation.

Fig. 2 shows an example of the spells created for a patent which was granted on 1/16/01 and cited three times on 6/14/01, 8/24/01, and 6/17/02.

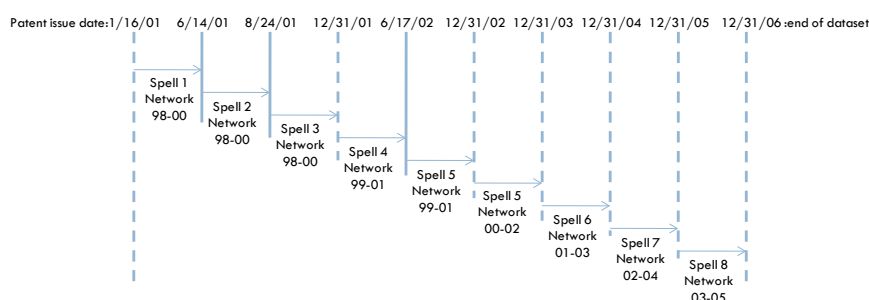


Fig. 2. Example spells for a patent in the dataset.

Note: The solid line denotes that the spell ended with a citation, the dashed line denotes that the spell ended with no citation. The next spell begins the day after the previous spell ends. The network indicates the time-span of the knowledge network that was used for computing measures for that spell.

The strategy of dividing time into spells effectively measures the effect of network measures of individuals who coauthored that patent on the citation of a patent through its entire life. The measures were computed on the basis of the network three years prior to the spell. That is, only inventors who had applied for patents in the three years prior to the spell were considered to be part of the network for that spell. This is in line with previous re-

search that shows that inventors are productive for three to five years (Rappa & Garud, 1992). We found support for this with the median productive life-span of an inventor being 3-5 years in our dataset.

3.4 Network measurement

In this sub-section, we describe the social network measures that were used to determine the status of inventors in the network. Based on previous studies (Nerkar & Paruchuri, 2005; Podolny & Stuart, 1995) three measures were selected for comparison: betweenness centrality, Bonacich power, Random walk centrality. We also proposed a new measure called Random Walk w/Teams which is likely to suit this problem domain more than other measures.

3.4.1 Betweenness centrality

This is a well-known and widely used betweenness measure proposed by Freeman (1979). Intuitively, BC of node k is defined as the fraction of times that a node i needs the node k in order to reach node j via the shortest path. BC for a node k is calculated as (Borgatti, 2005):

$$\sum_i \sum_{\substack{j \\ i \neq j \neq k}} \frac{g_{ikj}}{g_{ij}}$$

where, g_{ij} is the number of geodesic paths from i to j and g_{ikj} is the number of these geodesics that pass through k .

3.4.2 Bonacich power

The BP measure suggests that a node is important to the extent that it is connected to other important nodes. The importance of a node emerges recursively from the pattern of connections among all the inventors (this concept is similar to the PageRank (Brin & Page, 1998) algorithm). Details on the implementation of the measure can be found in Bonacich (1987).

3.4.3 Random walk betweenness (RW)

RW is a relatively new measure that includes contributions from all paths between nodes to calculate betweenness (Newman, 2005). RW for node k is equal to the number of times a random walk from i to j passes through k - averaged over all i and j . Thus, the measure includes paths that may not be optimal, though shorter paths still contribute more to the score. Details on the method can be found in Newman (2005). The measure also assumes that on each step during the random walk, information passes from the current node to one adjacent node (i.e., no parallel duplication). However, this assumption may not hold in knowledge networks of the kind studied here. Diffusion of information may happen in parallel within teams and follow a random walk outside them.

Random walk with teams (RWT)

Organizations generally have teams of inventors working together on projects. The communication levels within these project teams are much higher as compared to between teams (Hansen, 2002). We contend that there is close to parallel duplication of knowledge within teams, i.e., if one member of a team receives knowledge that is pertinent to the project, then all members of the team have access to it. With this assumption, we propose to add team identification to the RW measure to address the issue of parallel duplication.

Team identification

Fig. 3(a) shows a schematic of the assumed flow of knowledge within an organization with three teams. The circles in the figure indicate teams of inventors. The dashed arrows indicate parallel information duplication (within teams). The solid arrows indicate random walks between teams. As can be seen in the figure, we assume that knowledge diffuses in a parallel fashion within teams and flows through random walks outside them. In order to use this phenomenon to establish the status of inventors, we need to identify teams within an organization. There are several algorithms to identify communities or teams in social networks, Fortunato (2009) gives a comprehensive survey. Selecting an algorithm is a trade-off between computational times and accuracy; however, there have been few comprehensive comparisons of the algorithms on real-world networks. In addition, the choice of algorithms may also be defined by the network being studied; some algorithms may provide better results with certain kinds of networks. Our choice of an algorithm for this study was based on three factors: (1) Computational time was not a consideration since the networks were not large, (2) We wanted to select an algorithm that was known to be accurate enough to show that teams mattered in these networks, and (3) we needed an algorithm that had been used extensively on real-world intra-organizational networks. We selected the widely used community identification algorithm proposed by Girvan & Newman (2002) to identify teams in the collaboration network. The algorithm identifies cohesive communities using an iterative edge removal strategy based on betweenness measures. It has been shown to be superior to other community detection techniques (Newman & Girvan, 2004) specially in scientific collaboration networks. Another recent comparative study (Leon, Albert, Jordi, & Alex, 2005) showed that the algorithm had a satisfactorily high accuracy of close to 90%.

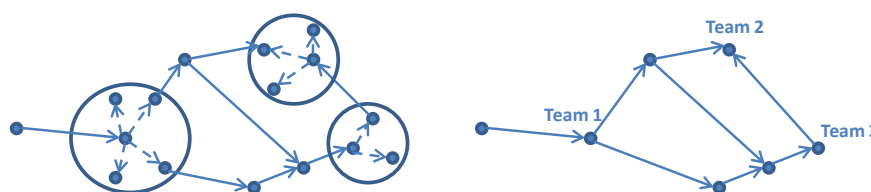


Fig. 3. (a) Flow of knowledge within an organization with teams (b) network after teams are identified and collapsed

Network collapse

Once teams are identified, the network is collapsed with each team replaced by a single composite node. Essentially, the composite ‘team’ node is structurally equivalent (Lorrain & White, 1971) to the combination of all the individual nodes in that it preserves the connections between the teams members and individuals outside. Thus, in Fig. 3(a) each of the large circles containing team members would be collapsed into a single node, the resulting nodes and connections are shown in Fig. 3(b).

Weak component identification

Once the network is collapsed, all the connected components in the network are identified and the measure is calculated for each component separately. This is different from some previous studies that use only the largest component in the network to calculate measures. This is important since different divisions of an organization may have self contained groups of inventors and calculating measures for the largest component would ignore smaller groups.

RW betweenness calculation

Random walk betweenness (using the Newman (2005) procedure for RW) scores are calculated for each node in each component of the collapsed network. Thus, the RWT measure calculated the RW betweenness score for entire teams taken as one node and single inventors who are not part of any teams. For statistical analysis, every individual in a team received the same RWT score. We believe that these new RWT scores will explain innovation diffusion better and identify individuals whose knowledge is valued for recombination within an organization.

3.5 Statistical analysis

We used patent citation data for statistical analysis since citation leaves a trail of how a patent builds upon previous innovations. Unlike in academic papers, patent citations are not likely to be superfluous (Singh, 2005; Sorenson, Rivkin, & Fleming, 2005). An intra-organizational citation of a patent is a choice made by the organization (and individuals within) to build on knowledge contained in the patent. In this study, we aim to ascertain if the network position of an inventor influences this selection process. Thus, the dependent variable is the citation of a patent by inventors other than those involved in its creation.

Cox proportional hazard models were used to study the effects of network measures on patent citation (other models including Weibull and Exponential were tested however, they were not found to be a good fit). The models used a repeated event hazard rate analysis to incorporate spells. These models were used since they incorporate both censored and uncensored cases, i.e., whether or not the patent was cited. Three kinds of variables were included in the statistical model: dependent variable: patent citation, explanatory variables: each of the social network measures, and control variables: factors (other than network measures) that effect patent citations. Since multiple inventors may be assignees on the same patent, a maximum of the social network scores among all the inventors for that patent was used as an independent variable.

Based on various previous studies, the following control variables were included:

- Calendar age: this controls for improvements in technology since the start of the dataset (Podolny & Stuart, 1995). As databases and information retrieval techniques improve, patents are easier to find and cite.
- Patent age: a patent is more likely to be cited if it has been around longer.
- Patent age squared: as the age of a patent increases, it may be outdated and less likely to be cited.
- Scope of a patent: The USPTO uses a technology classification system where a patent is classified into one or more technology classes. Studies have used the number of classes to represent the breadth of a patent that has an effect on the patent's impact (Lanjouw & Schankerman, 2001). We include this variable as the number of USPTO technological classes the patent is classified into.
- Number of claims: the number of claims indicates the value of a patent and the technological spaces it occupies or protects (Lanjouw & Schankerman, 2001).
- Age of prior art: Patents that build on old knowledge have different citation patterns than new ones (Nerkar, 2003). This is calculated as the median of the difference between grant year of the focal patent and that of the references cited within that patent.
- Self citation: A self-citation indicates confidence of an individual on his/her work. This may encourage other individuals to cite that work (Nerkar & Paruchuri, 2005). This is operationalized as a categorical variable which is a '1' if patent has been self-cited before spell, and '0' otherwise.
- Number of patent references/Number of academic references: Patents that cite more prior art may have a different influence than others. They may be in technologically crowded classes and have a different influence as compared to other patents (Fleming, 2001).
- Team size: One patent can have multiple inventors. When determining the effect of social network measures on the citation of a patent, we used the maximum of the measures among all the inventors of that patent. Including team size as a control variable accounts for effects of all inventors on the patent (Nerkar & Paruchuri, 2005) since a heterogeneity in team members can lead to differences in the influence of a patent (Reagans & Zuckerman, 2001).
- International presence of an inventor: Knowledge flows across international boundaries are different (Singh, 2007) and may affect the citation of a patent. This is operationalized as a variable that is set to '1' if any inventor on patent is outside the U.S. and '0' otherwise.
- Time to grant: A patent that is granted immediately may be uncontroversial and simple. A complex patent may take time to get approved. This might affect citation rates (Nerkar & Paruchuri, 2005).
- Technological effects: This controls for the difference in patenting across technological areas. Certain technological areas may cite a larger number of prior patents than others. This is operationalized as dummy variables for the top 20 classes (with ties retained) each organization patents in.

Based on the results obtained by previous node-level studies and the assumptions for knowledge flow in a network, four hypotheses were examined in this study with each in its own independent model. Each hypothesis tested the effect of an inventor's status (as established by a network measure) on the likelihood of his/her knowledge being selected by other inventors. These are summarized in Table 4.

Table 4. Hyptheses tested

	Measure	Effect
H1	Betweenness centrality	No effect
H2	Bonacich power	No effect
H3	Random walk	No effect
H4	Random walk w/Teams (proposed)	Positive effect

4. Experimental results and discussion

In this section, we show the results of the Cox proportional hazards analysis. Five models were constructed for IBM – one for control variables only (Model 0) and the others including one of the four measures (Model 1 – 4) along with the control variables. These models were constructed as described in Section 3.5. The correlation matrix (Table 5) shows that all correlations except those between some network measures are low and do not pose multi-collinearity problems. The high correlations between some network measures do not cause problems since each regression model contains only one measure.

Table 5. Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Random walk w/teams	1												
(2) Bonacich power	0.1040*	1											
(3) Random walk	0.4907*	0.2199*	1										
(4) Freeman betweenness	0.4094*	0.0084	0.7748*	1									
(5) Class scope	-0.0412*	-0.0364*	-0.0535*	-0.021	1								
(6) Prior age	-0.0673*	-0.0256	-0.0444*	-0.0167	0.0127	1							
(7) Patent refs.	0.0193	0.0506*	0.0319*	-0.0656*	0.0255	0.0698*	1						
(8) Academic refs	0.0745*	0.0104	0.0770*	0.1023*	-0.0523*	-0.0677*	0.0724*	1					
(9) Team size	0.1450*	0.1045*	0.2076*	-0.0365	-0.0296	-0.0529*	0.1958*	0.1209*	1				
(10) International presence	0.0592*	0.0123	0.0633*	0.1427*	-0.014	0.0553*	-0.2207*	0.0365*	-0.1036*	1			
(11) Time to grant	-0.022	0.0295	-0.0252	-0.0797*	-0.0370*	0.1610*	0.1936*	0.0178	0.1119*	0.0008	1		
(12) Claim count	0.0522*	0.0229	0.0573*	-0.0178	-0.0058	-0.0259	0.1016*	0.0905*	0.0590*	-0.0992*	0.0074	1	
(13) Self cited	0.1503*	0.0231	0.2581*	0.2434*	-0.0621*	0.0007	0.0278	0.1207*	0.1094*	0.0670*	-0.0415*	0.0285	1

Note: All correlation values above 0.05 are significant at $p < 0.05$ (*)

Table 6 shows the results for all four Cox regression models for IBM. The first column lists all the network measures and control variables. Each model (from Model 1 - Model 4) contained one network measure. As can be seen in Model 0, the likelihood of a patent being cited decreased (i.e., the hazard ratio < 1.0) with an increase in patent age and time to grant. This may be because as a patent increases in age, its contents become less relevant in a fast moving field like nanotechnology. The likelihood of patent citation increased with an increase in calendar age. The reason behind this may be the better availability of information retrieval technology and databases which make it easier to find a patent and cite it. The likelihood also increased with an increase in the claim count and academic references. As mentioned before, the claims are the number of ‘spaces’ occupied by the patent. More the spaces occupied, more likely the patent will be cited (Lanjouw & Schankerman, 2001). Self citation indicates, among other things, the confidence that an inventor has on his/her own patents. The model shows that the more an inventor self-cites the more likely others are to cite his/her patents. The significance of these control variables generally persisted across all models.

Model 1 shows that the BC score of inventors was found to have an insignificant effect on the citation of their patents. Thus, the measure does not adequately explain the effect inventor status on the selection of his/her knowledge for innovation. As discussed before, BC is based on the assumption that knowledge flows along shortest paths that may not suit this problem domain. Random walk (Newman, 2005) was also found to be insignificant (Model 2). This may be because even though the RW measure incorporates random diffusion and is a betweenness measure, it does not incorporate the influence of teams. Individuals between teams draw knowledge from diverse communities and the RW w/Teams measure is likely to perform better in this problem domain.

The Bonacich power of an inventor was found to be significant in Model 3. The measure has also been found to be significant by prior studies (Nerkar & Paruchuri, 2005). This implies that an inventor's knowledge is perceived to be more important (and cited) if he/she is connected to other important inventors. However, the absolute effect of the BP measure is very small since the hazard ratio is close to 1.0. A hazard ratio of 1.0 indicates that the variable does not increase or decrease the likelihood of a patent citation.

As can be seen from the table (Model 4), the random walk w/teams measure had a significant positive association with the citation of a patent. A unit increase in the RWT score of an inventor associated with a patent increases the likelihood of the patent being cited by 87%. This shows that the position of the inventor in a network positively effects the selection of his/her knowledge for recombination. There are three components to the RWT measure that may have contributed to its significance. Firstly, the focus on diversity of knowledge which implies that knowledge of inventors who have high betweenness scores is perceived to be valuable by an organization. Inventors with high betweenness are also likely to obtain knowledge from multiple disparate communities that may increase their innovative potential. Secondly, random diffusion is an important part of the RWT measure and this may have contributed to its positive significance. This is because information may not necessarily flow through shortest paths in a knowledge network (as shown by the insignificance of Freeman's betweenness centrality). A third factor is parallel diffusion, the RWT measure takes into consideration that knowledge can diffuse within a team from one individual to multiple individuals. These three assumptions in the RWT measure make it better suited to explain inventor status in the collaboration networks we study here.

Table 6. Cox regression results

	Model 0	Model 1	Model 2	Model 3	Model 4
Bet. centrality		1.3995			
Random walk			1.1211		
Bonacich power				0.9977**	
RW w/Teams					1.8700*
Patent age	0.9981**	0.9998	0.9998**	0.9998*	0.9998**
Calendar Age	1.3132*	1.5295	1.2942	1.4152	1.5933
Class scope	1 class*	1 class*	1 Class*	1 class*	1 class*
Prior age	0.9999	0.9997*	0.9999	0.9999	0.9999
Patent refs.	1.0012	1.0113	1.0012	1.0014	1.0015
Acad. refs.	1.0205*	1.0172**	1.0204*	1.0204*	1.0200*
Team size	0.9780	1.01567	0.975	0.9798	0.9695
International	1.2651	1.5459*	1.2593	1.2863	1.2544
Time to grant	0.9996**	1.0001	0.9996**	0.9996**	0.9996**
Claim count	1.0131**	1.0088	1.0131**	1.0129**	1.0128**
Self cited	1.5362*	1.8402*	1.5178*	1.5383*	1.4987*
Tech. effects	20 classes**	26 classes**	22 classes**	22 classes**	22 classes**

Note: All coefficient values above 0.05 are significant at $p < 0.05$ (*)

5. Conclusions

In this study, we examined the role of inventor status in knowledge networks on the selection of knowledge that is recombined to produce innovation in the nanotechnology field. A new network measure based on random walks and team identification (RWT) was proposed to model knowledge flow within an inventor collaboration network. Using empirical methods, it was found that inventor status as measured by RWT had a significant positive relationship with the likelihood that his/her knowledge would be selected for recombination. We believe

that the new measure in addition to modeling knowledge flow in a scientific collaboration network will help better understand how innovation evolves within organizations.

In future studies, we plan to test other important social network prestige measures like Burt's Structural Hole measures and information measures like flow centrality to test their effect on innovation selection and compare them to the proposed measure. In addition, we will examine the effects of collapsing the teams on other prestige measures. In doing so, we plan to conduct a similar study on multiple large organizations both individually and combined to a larger dataset to provide more validity to our results.

Acknowledgments This research was supported in part by "NanoMap: Mapping Nanotechnology Development," NSF, Grant #0533749 and the Faculty Development and Research Committee of Towson University. Portions of this paper were presented in the Hawaii International Conference on System Sciences (HICSS-42) in 2009.

References

- Agrawal, A., Cockburn, I. M., & McHale, J. (2006). Gone but not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships. *Journal of Economic Geography*, 6(571-591).
- Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, 45(3), 425-455.
- Albert, R., & Barabasi, A.-L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74(1).
- Bell, G. G. (2005). Clusters, Networks, and Firm Innovativeness. *Strategic Management Journal*, 26, 287-295.
- Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5), 1170-1182.
- Borgatti, S. P. (2005). Centrality and Network Flow. *Social Networks*, 27, 55-71.
- Breschi, S., & Lissoni, F. (forthcoming). "Cross-firm" Inventors and Social Networks: Local Knowledge Spillovers Revisited. *Annales d'Economie et de Statistique*.
- Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks*, 30(1-7), 107-117.
- Burt, R. (1992). *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
- Chen, H., Roco, M. C., Li, X., & Lin, Y. (2008). Trends in Nanotechnology Patents. *Nature Nanotechnology*, 3(3), 123-125.
- Chesbrough, H. W., & Tece, D. J. (1996). When is Virtual Virtuous? Organizing for Innovation. *Harvard Business Review*, 74(1), 65-74.
- Fleming, L. (2001). Recombinant Uncertainty in Technological Research. *Management Science*, 47(1), 117-132.
- Fleming, L., III, C. K., & Juda, A. (forthcoming). Small Worlds and Innovation. *Organization Science*.
- Fleming, L., King, C., & Juda, A. (2007). Small Worlds and Regional Innovation. *Organization Science*.
- Fortunato, S. (2009). Community detection in graphs (Publication no. arXiv:0906.0612v1). from arXiv.org:
- Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1, 215-239.
- Freeman, L. C., Borgatti, S. P., & White, D. R. (1991). Centrality in Valued Graphs: A Measure of Betweenness based on Network Flow. *Social Networks*, 13, 141-154.
- Friedkin, N. E. (1991). Theoretical Foundations for Centrality Measures. *American Journal of Sociology*, 96, 1478-1504.
- Girvan, M., & Newman, M. E. J. (2002). Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*, 99(12), 7821-7826.
- Hansen, M. T. (1999). The Search-transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly*, 44, 82-111.
- Hansen, M. T. (2002). Knowledge Networks: Explaining Effective Knowledge Sharing in Multiunit Companies. *Organization Science*, 13(3), 232-248.
- Henderson, R. M., & Cockburn, I. (1994). Measuring Competence: Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal*, 15, 63-84.
- Huang, Z., Chen, H., Chen, Z.-K., & Roco, M. C. (2004). International Nanotechnology Development in 2003: Country, Institution, and Technology Field Analysis based on USPTO Patent Database. *Journal of Nanoparticle Research*, 6, 325-354.

- Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.
- Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science*, 3(3), 383-397.
- Lanjouw, J. O., & Schankerman, M. (2001). Characteristics of Patent Litigation: A Window of Competition. *Journal Law and Economics*, 38, 463-495.
- Leon, D., Albert, D.-G., Jordi, D., & Alex, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics*(9), 8.
- Lorrain, F., & White, H. C. (1971). Structural Equivalence of Individuals in Social Networks. *Journal of Mathematical Sociology* 1(January), 49-80.
- Nerkar, A. (2003). Old is Gold? The Value of Temporal Exploration in the Creation of New Knowledge. *Management Science*, 49(2), 211-229.
- Nerkar, A., & Paruchuri, S. (2005). Evolution of R&D Capabilities: The Role of Knowledge Networks Within a Firm. *Management Science*, 51(5), 771-785.
- Newman, M. E. J. (2004). Who is the Best Connected Scientist? A Study of Scientific Coauthorship Networks. In E. Ben-Naim, H. Frauenfelder & Z. Toroczkai (Eds.), *Complex Networks* (pp. 337-370). Berlin: Springer.
- Newman, M. E. J. (2005). A Measure of Betweenness Centrality Based on Random Walks. *Social Networks*, 27, 39-54.
- Newman, M. E. J., & Girvan, M. (2004). Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69.
- Noh, J. D., & Rieger, H. (2002). Stability of Shortest Paths in Complex Networks with Random Edge Weights. *Physical Review E*, 66, 066127.
- Patrakosol, B., & Olson, D. L. (2007). How Interfirm Collaboration Benefits IT Innovation. *Information & Management*, 44, 53-62.
- Podolny, J. M., & Stuart, T. E. (1995). A Role-based Ecology of Technological Change. *The American Journal of Sociology*, 100(5), 1224-1260.
- Rappa, M. A., & Garud, R. (1992). Modeling Contribution Spans of Scientists in a Field: The Case of Cochlear Implants. *R&D Management*, 22(4), 337-348.
- Reagans, R., & Zuckerman, E. W. (2001). Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. *Organization Science*, 12(4), 502-517.
- Rosenkopf, L., & Nerkar, A. (2001). Beyond Local Search: Boundary-Spanning, Exploration and Impact in the Optical Disc Industry. *Strategic Management Journal*, 22, 287-306.
- Scott, J. (2000). *Social Network Analysis: A Handbook* (2 ed.). London, UK: Sage.
- Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, 51(5), 756-770.
- Singh, J. (2007). *External Collaboration, Social Networks and Knowledge Creation: Evidence from Scientific Publications*. Paper presented at the Danish Research Unit of Industrial Dynamics Summer Conference 2007, Denmark.
- Sorenson, O., Rivkin, J. W., & Fleming, L. (2005). *Complexity, Networks and Knowledge Flow*. Paper presented at the DRUID Tenth Anniversary Summer Conference, Copenhagen, Denmark.
- Sorenson, O., & Stuart, T. E. (2001). Syndication Networks and the Spatial Diffusion of Venture Capital Investments. *American Journal of Sociology*, 106, 1546-1588.
- Stephenson, K. A., & Zelen, M. (1989). Rethinking Centrality: Methods and Examples. *Social Networks*, 11, 1-37.
- Tsai, W. P. (2001). Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance. *Academy of Management Journal*, 41(4), 996-1004.
- Uzzi, B., & Spiro, J. (2005). Collaborations and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2), 447-504.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.