The Strength of Weak Ties as a Strategy to Allocate Research Funds: Making a Bioenergy Research Network More Productive

Syed Fuad, Texas Tech University, syed.m.fuad@ttu.edu

Michael Farmer, Texas Tech University, michael.farmer@ttu.edu

Selected Paper prepared for presentation at the 2023 Agricultural & Applied Economics Association Annual Meeting, Washington DC; July 23-25, 2023

The Strength of Weak Ties as a Strategy to Allocate Research Funds: Making a Bioenergy Research Network More Productive [1]

Abstract

The famous analysis of calling crickets proposed an answer to a conundrum that had puzzled biologists - how do crickets conform calls so quickly? A very small number of crickets (maybe one in 10,000) reacting to a distant call, in addition to an immediate neighbor could accelerate harmonization greatly. We test this small-world phenomenon in what is close to the universe of all publications on Black Liquor Gasification (BLG) at papermills. BLG can more than double electric power beyond the 2.1% of all U.S. electric output already sold by papermills from BLG boilers.

We collected the universe of published work from 1991 to 2007, author information, funding sources, and the number of publications. Using limited dependent variable methods, we estimate the number of publications and the entry or exit of active researchers within the network. We simulate each funding strategy over five cycles and update the network to create an outcome distribution.

Three funding strategies were compared for this experiment. All funding policies provide 22% more funding to the network overall. Direct Optimization funds those author pairs with highest number of expected publications, and on average, increases publications by 92% and researcher recruitment by 30%; the Smart Small World Rule funds author pairs that, once paired, maximize the number of other researchers who have collaborated with either coauthor, and increases publication rates by 113% and researcher recruitment by 32%; the Fairness Rule, which funds author pairs that reduce the average number of steps between any two researchers across the entire network, increases publication rates by 111% and researcher recruitment by 36%. Finally, providing no additional funding reduces publications and the number of researchers.

Overall, this experiment suggests that research funding that strengthens overall research community connectivity generates the highest levels of research productivity.

Keywords: network, black liquor gasification, small world, research funding, network simulation.

JEL codes: C45, D85, O31

[1] Parts of this work are drawn from a Doctoral candidacy defense on this topic.

1. Introduction

Social networks have been used to predict social outcomes in diverse aspects of social and economic relationships. The small world is a distinct property of social networks that explain, roughly speaking, how any two of nodes in the network can reach each other through a very short sequence of acquaintances (Kleinberg, 2001).

The effects of social networks have already been investigated in the fields of innovation and research and development (R&D). It has been found that the properties of a network have significant effects on stimulating new knowledge and innovation in a social network of researchers and innovators (see Watts, 2004; Hargadon, 2003; Cowan and Jonard, 2003; Baum et al., 2003; Schilling and Phelps, 2007; Uzzi and Spiro, 2006). Most of the discussed research work is descriptive, focusing on how an existing social network structures affect the realm of the scientific and academic networks (see Rumsey-Wairepo, 2006) or industrial innovation networks (see Schilling and Phelps, 2007; Fleming and Frenken, 2007; He and Fallah, 2009). The bulk of this scholarly descriptive work has focused on networks either of individuals or of institutions. This work seeks to employ specific properties of networks to locate strategic opportunities to inject funds into an existing research network to facilitate long run productivity and network resilience through time.

The core property that drives the small world phenomenon is the random long-range connection, especially if the distant nodes are tightly connected locally (Watts, 2004). We employ small world networks as well as other network structures to assess policy vehicles to improve research output in a social network of researchers in a specific field. We recognize that funding is a key input that drives the formation and evolution of a knowledge-based research network. This paper contributes to the research literature by providing an empirical study on the impact of funding injected strategically into a network to enhance its performance by facilitating specific network properties such as connectivity, density, degrees of separation, and path length – all related terms, but have specific technical definitions that differ.

We collected the universe of published work in the field of black liquor gasification from 1991 to 2007, alongside funding sources and amount, co-authorship details, and the number of publications. Currently, black liquor, a by-product in the pulping process of wood into paper, is boiled. This energy conversion process produces close to 2.1% of the U.S. domestic electricity supply. If black liquor could be gasified, this output could easily double, a substantial bioenergy contribution from a research activity involving relatively few researchers, mostly chemical scientists and engineers.

Using publication information, we represent all pairwise scientific collaboration as nodes in a social network and develop rules to guide the evolution of connectivity within the network. We use econometric techniques to obtain the probabilities of: expected publications (Poisson Regression); recruiting a new coauthor into the system (Multinomial Logit Regression); and exit from the network or connection break (Logit Regression) for each collaboration. Using these probabilities of success or failures for each connection, we simulate the expected number of publications, total authors, average shortest path length, and clustering coefficient for the whole network in the first period. Then we use the same econometric rules and simulate again in order to complete the second-period evolution process. In the same way, based on the new network after second period, we change the information of all variables for each connection, and perform the iteration three more times – a total of five times, each with 1,000 simulations. We then get a series of distributions of total publications, total authors, average shortest path length, and clustering coefficient for the network after five periods.

The funding strategy negotiation process involves three types of policy negotiation objects: Fairness Rule, where we fund author pairs with shortest average path lengths; Direct Optimization, where we fund author pairs with highest number of expected publications; and Smart Small World Rule, where we fund author pairs with highest number of first degree coauthors. In addition, we also see how the network evolves when no additional funds are injected. These are initial plausible policy choice rules that a policymaker can adopt and then compare the performance of these different options, including distribution of productivity, total authors, average shortest path length, and clustering coefficient, for the simulation exercises.

Our results suggest strategic funding helps to facilitate small world network formation. Networks grow compact over time and eventually exhibit short average path lengths and high clustering coefficients. Adhering to a policy that provides 22% more funding to the most efficient collaborators (Fairness Rule) increases publication rates by 111% and researcher recruitment by 36%. An equivalent funding to the most prolific researchers (Direct Optimization) increases publication rates by 92% and researcher recruitment by 30%; and funding the most prolific collaborators (Smart Small World) increases publication rates by 113% and researcher recruitment by 32%. Finally, providing no additional funding reduces publications and collaborations, but still marginally reduces average shortest path length but not clustering coefficient. Overall, we show that research funding strategies that strengthen overall research community connectivity appear to generate the greatest levels of research productivity.

2. Literature Review

An extensive amount of field data has been used through sharp mapping protocols to explain variation in labor market outcomes and production of creation in an economy. There is also extensive literature that explores how economic outcomes are influenced by social network structures. For example, studies by Boorman (1975) and Montgomery (1991) find a relationship between labor market outcomes and social network; Ellison and Fudenberg (1995) find that the structure of communication can affect a consumer's purchasing decisions; Bolton and Dewatripont (1994) show that organization of workers influence a firm's efficiency; in evolutionary game theory, Ellison (1993), Goyal and Janssen (1997) and Anderlini and Ianni (1996) all demonstrate the influence of network structure on possible coordination among agents.

As scientists seek to construct models of social process that result in observed structures of networks and study how the structures influence and facilitate the spread of knowledge, much of the interest in social networks revolves around understanding how networks develop and change (see Abrahamson and Rosenkopf 1997; Schilling and Phelps 2007; Doreian et al. 1996; Leenders 1996; Nakao and Romney 1993; Snijders 1998; Weesie and Flap 1990; Zeggelink et al., 1996). Moreover, many researchers like Cowan and Jonard (2001), Choi et al. (2010), and Granovetter (1973) have emphasized the significant effect of network topologies on the performance of the system and diffusion of innovation. Such dynamics analysis is important for understanding network stability and evolution, which in itself is necessary for understanding the effect of networks on individual and group behavior over periods. The clear importance of such problems has prompted a good deal of methodological research on network variation.

2.1 The Notion of Small-World and Six Degrees of Separation

The small-world notion states that any two people around the world who are randomly selected are connected to each other with some intermediate links. Milgram (1967) first conducted a quantitative survey regarding the small-world notion (Eslami, 2012) by randomly asking 296 individuals in Nebraska to deliver a letter to a specific person in Boston whom they did not know. His study concluded that each pair of people in the world is separated, on average, by six intermediate acquaintances. Later, this phenomenon was named "six degrees of separation", now more popularly known as the "six degrees of Kevin Bacon game". Watts and Strogatz (1998) later formalized this phenomenon by studying the collective synchronization of crickets. They introduced a model of small-world network in which there are some clusters that contain local ties among agents and also a few global links that enable connections between any pair of nodes in the network. More generally, this parameterized family of models exhibited an interesting combination of properties - high clustering and short pathlengths.

Watts and Strogatz (1998) also demonstrated the broad general application of weak ties to the sciences by studying the 'small world' phenomenon, which could be viewed in part as formal representations of the 'strength of weak ties'. These studies emphasize an important social characteristic central to many network theory applications: to accomplish a task in a large social network, agents need to navigate that network efficiently through weak ties to conduct, such as, a job search (Grannovetter, 1973). For example, workers are more likely to get the information of a new job through weak ties rather than strong ones. Sometimes the most useful resource is obtained from the occasional person whose connection is outside of the local associations.

Many empirical studies identify the small-world property operating in diverse and vibrant social networks. These include networks of American corporate boards, German corporate industries, strategic alliances, Canadian investment bank syndicates, email networks, Italian scientific and academic collaboration networks, and invisible scientific colleges (see Kogut and Walker 2001, Davis et al. 2003, Verspagen and Duysters 2003, Baum et al. 2003, Dodds et al. 2003, Balconi et al. 2004). Powell (1990) and Liebeskind et al. (1996) show that the property of network ties that describes a network among existing among individual researchers and inventors, housed at different universities and companies are among the most important factors that move technology forward to have a significant effect on the knowledge productivity.

It is not surprising that small-world researchers have proposed that the small-world network structure in a research social network would have an immense effect on the production of knowledge and innovation (see Watts 1999; Hargadon 2003; Cowan and Jonard 2004; Baum et al.

2003; Schilling and Phelps 2007; Uzzi and Spiro 2005). It has been hypothesized that a network that exhibits a strong small-world property can enhance information transmission efficiency among the network actors. We suggest using funding as a strategy to directly install small world property into a network by funding specific collaborations rather than specific researchers.

Our objective is to assist the policymaker in the search for strategic connections that will facilitate greater research output. We suggest that the individual node (person or institution) may be a more efficient way to use the rich data available from modern social network maps. To this end, we suggest that there is a series of dynamic methods that can be applicable to certain types of networks, yet still honor the basic small world insight of long-range ties. When public administrators wish to encourage specific ties as a mechanism to facilitate high tech job creation or to stimulate research, encourage investment, or mobilize more vigorous community action, the ever-increasing quality of social network mappings, we contend, contains more information than many current techniques utilize. As applications of social networks in this field grow in sophistication through advanced data collection and mapping tools, construction of more developed instruments is required to entirely use all of this data, especially individual's information regarding how they acquire resources and accomplish specific tasks.

3. Research Questions

This paper addresses two main research questions. First, we examine to what extent the structure of the collaboration network of black liquor gasification scientists resembles the small-world network structure. Since this type of network structure has attracted interest from researchers and has been shown to facilitate knowledge creation and diffusion, it is important to determine if the network shows small-world characteristics. To do this, we provide statistical evidence to evaluate the role of network structure by quantifying the properties of scientists' network.

Through this, we ask if the small-world structure facilitates the knowledge creation and the innovative performance of the inventors in the field of black liquor gasification. Previously, it has been accepted by scholars (see Cowan and Jonard 2004, Schiling and Phelps 2007, Watts 1999, Hargadon 2003) that the small world structure enhances the innovative productivity of the inventors' network to a considerable extent.

Second, we examine the effect of the different funding strategies on the evolution of the structure of collaboration network of scientists in the field of black liquor gasification on their research productivity and connectivity. We are interested in testing various structural properties of the network and assessing the impact of funding on the creation of knowledge by scientists.

Furthermore, the bulk of prior research analyzed the effect of patent coinventorship networks on the innovation productivity of inventors; but this study takes one step further by taking into account the important role of scientists' reciprocal knowledge transfers during the creation of their scientific knowledge (represented here by the article coauthorships), in promoting the innovativeness of scientists.

3.1 Why Black Liquor Gasification?

We are interested in this field because black liquor gasification offers significant improvements in energy efficiency and environmental performance, as well as economic benefits. Currently, black liquor, a by-product in the pulping process of wood into paper, is boiled. This energy conversion process produces close to 2.1% of the US domestic electricity supply. If black liquor could be gasified, this output could easily double, a substantial bioenergy contribution from this one research activity involving relatively few researchers, mostly chemical scientists and engineers.

This small network has the benefit that it allows us to collect an almost exhaustive list of researchers, their publications, authorship, citations, funding, funding sources, etc. that ultimately allows us to represent all collaborations as a social network to compare network evolution under different funding strategies. Moreover, the boilers, most of which were established in the 70s and 80s and need to be replaced soon. Therefore, continued strategic funding is needed in this area to stimulate research, innovation, and development and ultimately contribute to a larger contribution to environmental-friendly energy sources.

4. Data Description and Methods

The data set in this study is based on a survey of 20 top researchers in black liquor gasification, collectively responsible for the bulk of all publications in this topic area. From these surveys, acquired information on 126 publications, published in 40 journals by 127 researchers from 60 institutions. From each author's data in the survey, we extract information on the date of publication, the names of all coauthors, the presence of any new authors entering the network in a particular paper, and the funding level of the corresponding project.

Next, we assess how different factors affect the probabilities of success (publication productivity and recruiting a new researcher) and failure (exit of a researcher) for any author-pair connection. Funding is one independent variable that explains successes. Drawing from Wang (2009), we also construct several new variables using the original data based on the author-pair connection unit.

1. Dependent Variable: number of papers published by any two authors (e.g., if authors A and B published 5 papers together, then y is equal to 5)
2. Independent Variable - Funding: total funding level for any two authors' publications. Suppose author A published 15 papers and author B published 10 papers. Suppose that a total of 20 unique papers (not including the 5 repeated papers) are attributed to 2 projects. m denotes the funding of project 1 and n denotes the funding of project 2, then $(m+n)/20$ will be the value of the variable – funded dollars per publication.
3. Independent Variable – First degree: total first degree coauthor relations in the network for any two authors. Suppose author A is directly connected with 5 other authors, and author B is directly connected with 6 other authors, then the value of this variable is equal to 11. If there are duplications, only unique coauthors are counted.
4. Independent Variable – Second degree: total second degree relations, or those two degrees of separation from a author-pair. We define second degree relations as follows: if author A is only indirectly connected with author C or author D through her coauthorship with B, then A is defined as being in a second degree relation to C and D. Total persons two degrees away count as the number of unique 'coauthors of coauthors'.
5. Independent Variable – Shared coauthors: number of total coauthors from any two authors' shared publications. Suppose author A and author B published 5 papers together, and these 5 papers involve 15 authors, then the value of this variable is equal to 15. This therefore is the number of other coauthors a particular author-pair shares directly in their joint work.
6. Independent Variable – New authors: total number of new authors introduced to the network in publication by a coauthor-pair unit from any two authors' publications (e.g., if authors A and B published 10 papers together, and there are 4 new authors in these 10 papers, then the value of this variable is equal to 4).

7. Independent Variable – Funding square: the square of the Funding variable.
8. Independent Variable – First degree x second degree: the interaction between First degree and Second degree coauthors.

4.1 Methodology

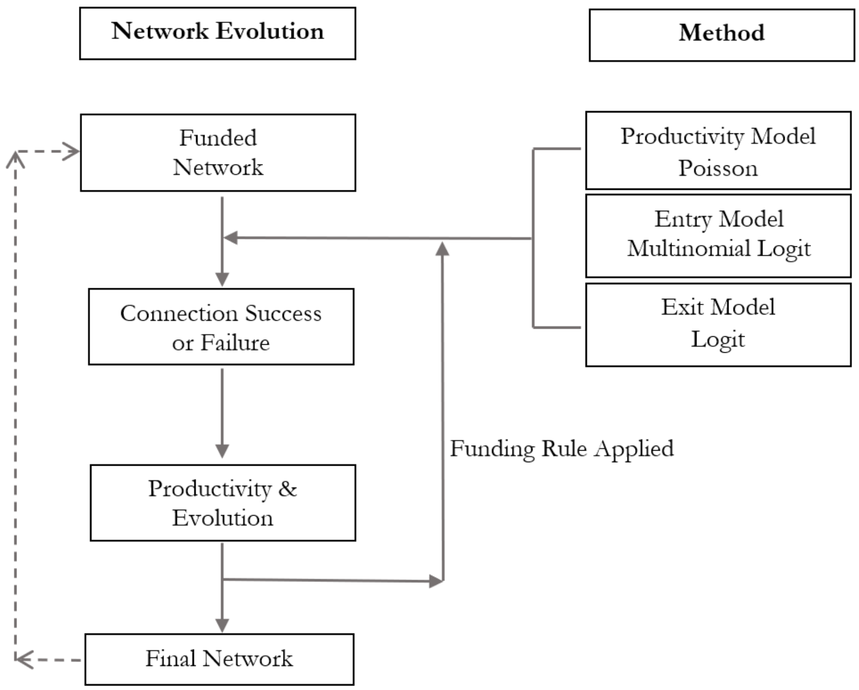
4.1.1 Evolution of the Network

To study the evolution of the connectivity within the network, we use the following rules (graphically illustrated in Figure 1):

- A Poisson regression is used to estimate the productivity for a connection. With the probability from the regression, we can simulate the number of publications for the whole network.
- A Multinomial Logit regression is used to estimate the probability of a new permanent entrance into the network; with the probability of entry for each pair of authors we can simulate how many new authors will enter the network.
- A Logit regression is used to estimate how many researchers permanently exit a network. First, we find the probability of whether a connection is broken or not. Second, if this connection is broken, we find which author in the connection exits the system. Then with the probability of exit for each pair of cells we can simulate how many authors will exit from the network.

The starting network is composed of 127 researchers who published 126 papers in black liquor gasification. If any two authors have at least one publication, we connect the pair as collaborations, and extract the dependent and independent variables: how often they publish, how much funding they receive, whether their shared publications brought in new researchers in the network (and how many) and whether they failed to retain researchers already in the network (and how many).

Figure 1: Network Evolution



4.1.2 Funding Strategies Guiding Evolution of Network

The funding strategy negotiation process involves three types of policy negotiation objects: Fairness Rule, where we fund author pairs with smallest path lengths; Direct Optimization, where we fund author pairs with highest expected publications; and Smart Small World Rule, where we fund authors with highest number of first degree coauthors. These are initial plausible policy choice rules that a policymaker can adopt and then compare the performance of these different options, including distribution of productivity, total authors, average path length, clustering coefficient and density analysis, for the simulation exercises.

5. Results

5.1 Parameter Estimation

5.1.1 Productivity Model

The first model for observing the productivity of a network is estimated by Poisson regression process. The unit of analysis is each connection or pair of researchers, and the dependent variable is the number of publications. Since the number of publications is discrete units with zero being the largest number, a Poisson estimator is used to predict the expected number of publications by a collaboration pairing. The method also provides a discrete probability of publishing a given number of papers, which can subsequently be employed in randomized simulation.

The results are given in Table 1, which contains the coefficient estimates. The independent variables are funding, which is total funding level for any two authors' publications; first degree, which is total first-degree connection in the network for any two authors; second degree, which is total second-degree relationships, or those two degrees of separation from a co-authoring pair; shared coauthors, which is number of total coauthors from any two authors' shared publications; new authors, which is total number of new authors introduced to the network in publication by a collaborating unit from any two authors' publications; funding square, which is the square of the independent variable of funding; and finally, first by second degrees, which is the multiplication between the second independent variable and the third independent variable.

The estimated regression results indicate the following. In most cases, the independent variables are significant with p-value less than 0.05, except variables for second degree and for new authors. We are primarily interested in the effects of funding, so a discussion of the regression results will be limited to that variable, and first and second degree relationships, although as subsequent tables reveal, most other variables are also significantly related to the dependent variables.

The effect of funding level on publications is very significantly different from zero, and the result means that the new average expected publication will be equal to the old average publication multiplied by $e^{0.236}(pubs = pubs * e^{\beta_1})$ when the funding level is increased by one unit or average probability that any two authors will successfully publish their average number of work increases by 23.6% as the research funding level increases one unit, measured in $100,000 units.

First steps relationship is also an important variable affecting the publication probability. In Table 1, the probability that pairs of researchers will publish increases by 7.5% for every additional unit of first degree relations, and the new average expected publication will be equal to the previous average publications multiplied by $e^{0.075}(pubs = pubs * e^{\beta_2})$.

By the same criterion, the variable of coauthors affects the probability of success for each collaboration in a positive way. The greater the number of shared coauthors a pair of researchers enjoys, the probability of publication success improves by 44.7%. The variable of funding square is negative, which shows a nonlinear relation with the probability of success.

What is striking in the result is the power of additional co-authorship, where an additional co-author increases publication success by 44.7% as a result the new publication will be equal to old average publications multiplied by $e^{0.447}(pubs = pubs * e^{\beta_4})$. This leaves open the policy concern addressed below regarding the relative strength of different resources: direct funding as a resource, or collaboration capacity and overall connectivity. Results above suggest both are strong. From a policy perspective, given the expense of collecting funding and past project information versus generating a research map, results suggest that network position of co-authorship history might be a source for a viable stand-alone policy.

As indicated above, the primary interest in estimating the model is to obtain the estimates of the independent variables which affect the probability of success for any pair of researchers. With the parameters, we calculate the expected of publications of each pair of authors, and we could estimate the probability that any collaborations could publish certain number of papers $P(x = h), h = 0, 1, 2, \dots, 20)$. These probabilities allow us to simulate the total number of publications for the whole network.

5.1.2 Entry Model

The second model for observing the success of recruiting a new person into the system is estimated by multinomial logit regression. The unit of analysis is also each pair of researchers.

The multinomial regression gives us the parameters estimates for entry model which is used to get the probability that any pairs of researchers could make a new person enter the system. The results are illustrated in Table 2. Here, group 1 indicates coauthor pairs who do not bring in additional authors and group 2 indicates coauthor pairs who bring in additional authors who only continue once. Because we only care about how many new researchers enter the system permanently, we focus on analyzing the probability of group 3.

The effect of variables for funding, first steps, number of shared coauthors, and new authors on probability of entry for group 3 which denotes the case that this collaboration could make new people come into the system and the new author will not exit the system.

Once we have the estimated probabilities for each pair of authors in our data set to make a new person enter the system and they will stay in the system and continue to publish in future $\left(P(x=3)\right)$. These probabilities allow us to simulate the total number of new authors for the whole network in section 4 for static simulation.

5.1.3 Exit Model

The third model is used to estimate the probability of exit for each pair researchers by a logit regression. The dependent variable is either zero, meaning the connection will not break, or one meaning that this connectivity will break. There are two steps to complete this objective. First, we estimate if the connection for a combination of authors breaks or not, and second if it breaks, which author of the two will exit the network. We calculate the probability for exit for each researcher and compare each's probability of exit to determine which of the two will leave.

The logit regression for exit is shown in Table 3. Most independent variables are significant at 5% level. Most variables that are statistically significant have a negative effect on probability that the connections will break for each pair; the results suggest increasing funding level, first and second degree connections, and shared coauthors will on average decrease the probability of exit for each collaboration, which makes intuitive sense.

5.2 Static Baseline Simulation

We start off by simulating the baseline network over one period to illustrate the simulation rules and principles. Based on the original data set, we first run the three regression models once and simulate the number of publications (Poisson Regression), new author (node) entry (Multinomial Logit Regression), and author exit (Logit Regression).

5.2.1 Poisson Simulation

We first calculate $(P(x=h), h=0,1,2,\ldots,19,20)$ where h is the number of actual publications. Based on the parameter lambda for each connection, we are able to generate a random number drawn from the Poisson distribution corresponding to the parameter for each collaboration. We do this for all observations (author connections) in the data set, and then sum all simulated numbers and obtain a total connection of 688 for the whole network. This process is repeated 1,000 times to eventually obtain a distribution of total connections.

An important point to note is that this process has a duplication problem because this is a pairwise author's estimation process. For example, we get 688 expected publications in the first iteration, but each publication may include more than a pair of coauthors. Hence, we have to scale by the ratio of real number of papers to real connections to get an estimated number of publications.

$$\frac{real\ connections}{real\ publications} = \frac{estimated\ connections}{estimated\ publications} => \frac{682}{126} = \frac{688}{x} => x = 126$$

where x is the estimated number of publications in the static baseline case.

5.2.2 Multinomial Logit Simulation

Next, we simulate the number of new author entry into the network. We first calculate the probability of recruiting a new person for each combination (y=3; where 3 indicates whether the author pair recruits a new permanent author), and then draw a uniform random number to compare with the estimated probability. If the random number is larger than the estimated probability, then this coauthor pair/connection does not recruit a new coauthor, and vice versa. Again, using 1,000 iterations of the simulation, we find that around 136 new connections on average will enter the network in one funding cycle. However, the same duplication problem exists, and can be similarly adjusted to estimate 25 new people entering the network.

$$\frac{real\ connections}{real\ number\ of\ authors} = \frac{estimated\ connections}{estimated\ new\ authors} => \frac{682}{127} = \frac{136}{x} => x = 25$$

5.2.3 Logit Simulation

Finally, we simulate the number of author-exits from the network. To do this, we first calculate the probability that a connection is broken (y=1, where 1 indicates whether one of the authors from a connection exits the network), and use a uniform random number to compare with the estimated probabilities. If the random number is lower than the probability, then this connection is predicted to break. We also calculate each single author's probability of exit; so, when we identify a connection involving that author is broken, we are able to compare the two authors' individual probability of exit and decide which author will exit the system. After 1,000 simulations, we conclude that about 38 authors exit the system. Coupled with 25 new authors entering the network, an exit of 38 authors implies 13 author exits from the network and a total of 126 papers being published in the static baseline scenario.

5.3 Stimulative Dynamic Funding Policies

Static simulation can only give one possible outcome in one period, that too without funding; so we do not know how a network evolves in response to different funding strategies. The funding policy negotiation process involves three types of policies: Fairness Rule, Direct Optimization, and Smart Small World Rule. These initial plausible policy choice rules allow a policymaker to compare the distributions of productivity, active researchers, and average pathlength, all of which we use to assess the network properties under different funding scenarios.

5.3.1 Fairness Rule

This funding strategy funds author pairs with the shortest average pathlength and is based on the principle that linking clusters increases network activity. The selected process seeks to link collaborations that minimize shortest pathlength among nodes in the network. Mathematically, let d_{ij} denote the shortest distance between authors i and j, then the average length is minimized as:

$$Min \ \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

where n is the number of vertices in the network.

Simply put, connections are chosen so that the average number of steps along the shortest paths for all pairs of network nodes is as small as possible. Generally, there are many connections to choose from (10-15 connections that typically have the same smallest number pathlengths), so in reality, the policymaker is not limited to a small set of options. In our case, we choose 5 connections (pairs of authors) with the minimum shortest pathlength and give each $100,000, connect them and iterate the simulation five (5) times. This implies $500,000 of additional capital in each funding round in total and $2,500,000 total in five rounds of funding, equivalent to 22% of the total funding in the data we have. Figure 2 demonstrates the progressive evolution of the network across baseline, first, third and fifth funding rounds. Each node represents an author, the vertice/edge connecting any two nodes indicates whether the nodes published together, and the width of the vertice indicates the strength of a relation.

The distributions of publications for both baseline simulation and smart small world are shown in Panel A in Figure 3. We observe a significant increase in publications after applying funding the most efficient collaborators; the mean number of publications increases by 111% to 266 publications compared to the Static Baseline case of 126 publications, which is a significant increase in productivity in response to 22% more funding over five cycles.

The distributions of the number of authors and average pathlength for Static Baseline are shown in Panels B and C of Figure 3. The mean of the number of authors in Static Baseline is 114, while with additional stimulative funding increases the mean to 155, equivalent to a 36% increase.

The third measure we evaluate is the shortest pathlength, also called geodesic distance, which is defined as the path with the minimum number of edges measuring the shortest possible distance between any two nodes; the shortest pathlength of a network is thus simply the average of the shortest pathlengths of all nodes. A typical small world property is small shortest pathlength. A large network size does not necessarily occlude formation of small world networks, as such networks are often large yet still exhibit short pathlengths and compactness (Albert and Barabasi, 2002). In this case, the mean of shortest pathlength reduces significantly from 3.898 in the Static Baseline to 2.897 in response to funding, which suggests that following this policy may stimulate authors to build more first degree connections in addition to publishing more.

Overall, we find that this policy leverages the properties of efficient collaborators and selects new research collaborations based on their contribution to overall network connectivity and leveraging this strategy will generate 111% more publications and recruit 36% more researchers to the network with only 22% more funding. The results also suggest that it is possible to inject funds strategically into research networks and infuse in them properties typical of small world networks, and doing so helps increase research productivity and researcher connectivity and overall network efficiency.

5.3.2 Smart Small World Rule

The second stimulative policy selection is Smart Small World Rule. This selection process links those collaborations that maximize the total number of connections in the network and works by maximizing new authors one degree of separation away from a combination or research collaboration. The rule tends to draw links between highly connected points to highly connected points on the other. Formally, the rule maximizes C_{ij}:

$$Max \sum C_{i,k} + \sum C_{j,m} + C_{ij}$$

Simply, connections are chosen such that the total number of final connections (total first degree connections for two authors) that attach each of the two authors (nodes) linked is as large as possible. Like the last policy, we choose 5 pairs of researchers with the maximum first degree connections in the network. Each of the pairs is provided with $100,000, connect them and run the same process five times. Figure 4 demonstrates the evolution of the network in response to this funding policy.

The distributions of publications for both Static Baseline and Smart Small World Rule are shown in Panel A of Figure 5. We see a significant increase in the number of publications after applying funding over five cycles; the mean of the number of publications for Smart Small World Rule increases by 113% to 268 publications compared to the Static Baseline case of 126 publications.

The distributions of the number of authors and shortest pathlength for Static Baseline are shown in Panels B and C of Figure 5. The mean of total authors in Static Baseline simulation is 114, while providing additional funding increases this to 150, equivalent to a 32% increase.

The mean of shortest pathlength also reduces from 3.898 to 2.992, which suggests that funding the most prolific collaborators will increase publications, author connectivity and make the research network more compact and efficient, all desirable properties of a small world network.

5.3.3 Direct Optimization

A third policy criterion is the Direct Optimization Rule. If the policy makers' goal is publications, then it makes sense to put two highly published authors together to increase overall publications. This rule locates the greatest increase in expected output from single a connection. It uses information beyond connectivity, based on expected collaboration success, in this case, collaboration success means the publications for any connection.

We define a collaboration success as S_{ij} (expected numbers of publications) for any two nodes $i,$ and j connected by vertice C_{ij}. We define the probability of success for that collaborating in any period as $P(S_{ij})$. If a given vector of node and collaboration specific characteristics, x, affects the probability of success on a given task for a collaboration, then we define the function $P\left(S_{ij}\right) = S(C_{ij}(x))$ to present the probability of success over a distribution. Here, x are characteristics such as the number of total connections that a node has and the history of past successes for that collaboration, C_{ij}. Technically, the Direct Optimization rule searches for a collaboration, C_{ij}, to obey the objective (maximize the probability of expected publications):

$$Max \sum P(S_{ik}) + \sum P(S_{jm}) + P(S_{ij})$$

The Direct Optimization Rule differs from the Smart Small World rule which maximizes the number of connections between the most highly connected nodes, emphasizing instead connections that contribute the largest expected marginal gain. Measuring the instantaneous output gain for the two nodes linked and spillover impacts on all of their connections, the rule chooses those connections that are emergent productive collaborations that have not fully matured.

Like the previous policies, we again choose 5 pairs of researchers, who have the maximum expected publications in the network and assign $100,000 to each pair as in the original case, connect them and run the same process over five periods. Figure 6 demonstrates the gradual evolution of the network under this funding policy of funding the most productive researcher pairs.

The distributions of publications for both static baseline simulation and Direct Optimization are shown in Panel A of Figure 7. The result shows a significant increase in publications after funding author collaborations with the highest expected publications. The mean of publications for Direct Optimization increases from 126 to 242, equivalent to a 92% increase, which means that this policy too could have a stimulative effect on the productivity of network. Surprisingly though, even though this policy, by design, funds the most prolific researcher pairs, compared to the Direct Optimization and Smart Small World Rule, this policy resulted in a smallest increase in publications.

The mean of total number of authors in Static Baseline is 114, while with Direct Optimization policy the mean of total authors increases 30% to 148 (Panel B of Figure 7), which shows an increase in author recruitment in Direct Optimization compared to Static Baseline, although again, this increase in the number of authors is smaller compared to Fairness and Smart Small World Rules. Finally, the shortest pathlength however, decreases the most compared to the other funding policies; from 3.898 to 2.952 (Panel C of Figure 7).

Overall, funding the most prolific authors yields more publications and authors and enhances network connectivity; however, the increase in the number of publications and number of authors is lower relative to the other two policies.

5.3.4 No Funding

Finally, providing no funding to any coauthor pair and iterating the simulation five times results in the evolution of the network as shown in Figure 8; we observe a marginally smaller number of publication outputs (126 to 116; an 8% decrease as shown in Panel A of Figure 9), significantly smaller number of authors (114 to 90; a 21% decrease as shown in Panel B of Figure 9), and interestingly, a reduction in shortest pathlength (from 3.898 to 2.992 as shown in Panel C of Figure 9).

5.3.5 Policy Summary and Network Evolution

We suggest three policy tools based on different goals. The Fairness Rule funds those author pairs with the shortest pathlength, and yields the highest number of authors after five periods of funding simulation. The Smart Small World Rule funds author pairs with the highest number of first degree connections, and yields the highest number of expected publications. Direct Optimization funds author pairs with the highest number of expected publications and produces the smallest increase in the number of expected publications and author recruitment equivalent to the Smart Small World rule.

All funding policies yield shorter average path length compared to the Static Baseline; one reason why that happens is demonstrated by the generation of large clusters in the graphs. In all funding policies, as more authors enter the network, initially, they are connected only to the coauthors by whom they were recruited. As the funding cycle progresses, these new authors begin to form connections with other authors outside those who initially recruited them, resulting in the growth of multiple interconnected clusters, which expectedly decreases the length of connectivity for these new authors to others in the network. The results also suggest that it is possible to inject funds strategically into research networks and infuse in them properties typical of small world networks, and doing so helps increase research productivity and researcher connectivity.

Table 4 presents the summary diagnostics of the simulations. The table also presents two additional measures of network clustering: clustering coefficient and small world index, both measures used to evaluate small world properties in networks. The small world index is a calculated variable measured by dividing the clustering coefficient by the shortest pathlength (Eslami, Ebadi, and Schiffauerova, 2013). Generally, the higher the clustering coefficient and small world index, the more dense and compact the network is, and in our context, the more closely connected and efficient are the researchers in the network. Again, under all funding policies, the clustering coefficient and density are higher than the static baseline case.

Table 4: Policy Table

	Avg. # of publications	Avg. # of authors	Avg. shortest pathlength	Clustering coefficient	Small world index
Static Baseline	126	114	3.898	0.442	0.113
Fairness Rule	266	155	2.897	0.499	0.172
Smart Small World Rule	268	150	3.073	0.500	0.163
Direct Optimization	242	148	2.952	0.480	0.163
No Funding	116	90	2.992	0.357	0.119

One important thing to note is that the networks under different funding scenarios cannot always reliably be compared with each other if the networks have significantly different size or connections. The level of entry and exit from network can potentially make a network appear less successful (more researchers with few coauthors, higher shortest pathlength), or more successful by showing greater density by eliminating the loose link authors to other groups (severing weak connections that remove a whole group from the network). Therefore, the relative success of a funding strategy should be assessed by considering the performance of several properties including the number of new author entries, exits, publications, and network connectivity measures.

6. Conclusion

We propose to use the small world property strategically to direct funding to specific research collaborations that enhance small world properties in a research network. As an experiment, we utilize the actual record of publication output of a network of chemical engineers who had worked on a bioenergy technology. We were able to attach specific grant awards to specific publications, and to impacts of those collaborations on bringing in (or losing) researchers to the network. We simulate future outcomes under alternative funding strategies among those same researchers in that research network to facilitate research productivity, connectivity, and resilience over time.

For this experiment we collected the universe of published work in the field of black liquor gasification from 1991 to 2007. The researchers identified for us the funding sources, the size of awards and coauthorship details for each publication. Currently, black liquor, a by-product of paper pulping, is processed in a boiler. This energy conversion process at the time produced close to 2.1% of the U.S. domestic electricity supply. If black liquor were processed in a gasifier instead, power output could have doubled, a substantial contribution from a singular research activity among relatively few researchers.

Using this publication information, we construct a network. Each researcher is a network node, and each edge (connection line) connects two researchers who have coauthored one or more papers. The network evolves from coauthor successes or failures from its baseline by a series of stochastic events: the probability of the number of publications between coauthors based on funding level (including no funding); the probability of a given coauthorship pair recruiting new researchers; or the probability that a given researcher (node) falls away from the network. These outcomes are stimulated by new funding, yet depend on the past successes of collaborators, which itself is a product of the number of collaborators involved and the productivity of those collaborators. These rules direct the evolution of specific connections between network researchers, which are altered by fund decisions.

We use econometric techniques to obtain the probabilities of: the expected number of publications (Poisson regression); no recruitment of new coauthors, or the recruitment of new coauthors who publish once, or publish two or more works (Multinomial Logit regression); and the chance that a given researcher may exit the network (Logit regression). Drawing from the distribution of these probabilities, we simulate network changes in the expected number of publications, the total number of network participants, the overall average shortest path length between network members, and a network clustering coefficient. We then repeat the process to simulate evolution of the second period on this updated network, and so on for five periods. We run each funding rule 1,000 times through each of the five period treatments (a total of 5,000 times), drawing from the probability distributions for each type of outcome for each connection. For any treatment strategy, we obtain distributions of end period outcome: total publications, total number of researchers, the average shortest path length between any two authors in the network, and a network clustering coefficient.

Three funding strategies were compared for this experiment. Direct Optimization funds those author pairs with highest number of expected publications; a Smart Small World Rule funds author pairs that, once paired, maximize the number of other researchers who have collaborated with either coauthor; and, finally a Fairness Rule that funds author pairs that reduce the average number of steps between any two researchers across the entire network.

All funding policies provide 22% more funding to the network overall. Direct Optimization, on average, increases publications by 92% and researcher recruitment by 30%; the Smart Small World Rule increases publication rates by 113% and researcher recruitment by 32%; the Fairness Rule increases publication rates by 111% and researcher recruitment by 36%. Finally, providing no additional funding reduces publications and the number of researchers.

Overall, this experiment suggests that research funding that strengthens overall research community connectivity generates the highest levels of research productivity, using the properties of success in a real world research network that contained the universe of research funding and research output by those researchers.

References:

1. Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. Organization science, 8(3), 289-309.
2. Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 47.
3. Anderlini, L., & Ianni, A. (1996). Path dependence and learning from neighbors. Games and economic behavior, 13(2), 141-177.
4. Balconi, M., Breschi, S., & Lissoni, F. (2004). Networks of inventors and the role of academia: an exploration of Italian patent data. Research policy, 33(1), 127-145.
5. Baum, J. A., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from?. Industrial and Corporate change, 12(4), 697-725.
6. Bolton, P., & Dewatripont, M. (1994). The firm as a communication network. The Quarterly Journal of Economics, 109(4), 809-839.
7. Boorman, S. A. (1975). A combinatiorial optimization model for transmission of job information through contact networks. The bell journal of economics, 216-249.
8. Choi, H., Kim, S. H., & Lee, J. (2010). Role of network structure and network effects in diffusion of innovations. Industrial marketing management, 39(1), 170-177.
9. Cowan, R., & Jonard, N. (2001). Knowledge creation, knowledge diffusion and network structure. In Economics with heterogeneous interacting agents (pp. 327-343). Springer, Berlin, Heidelberg.
10. Cowan, R., & Jonard, N. (2003). The dynamics of collective invention. Journal of Economic Behavior & Organization, 52(4), 513-532.
11. Davis, G. F., Yoo, M., & Baker, W. E. (2003). The small world of the American corporate elite, 1982-2001. Strategic organization, 1(3), 301-326.
12. Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An experimental study of search in global social networks. science, 301(5634), 827-829.
13. Doreian, P., & Mrvar, A. (1996). A partitioning approach to structural balance. Social networks, 18(2), 149-168.
14. Ellison, G. (1993). Learning, local interaction, and coordination. Econometrica: Journal of the Econometric Society, 1047-1071.
15. Eslami, H., Ebadi, A., & Schiffauerova, A. (2013). Effect of collaboration network structure on knowledge creation and technological performance: the case of biotechnology in Canada. Scientometrics, 97, 99-119.
16. Eslami, H., Ebadi, A., & Schiffauerova, A. (2013). Effect of collaboration network structure on knowledge creation and technological performance: the case of biotechnology in Canada. Scientometrics, 97(1), 99-119.
17. Fleming, L., & Frenken, K. (2007). The evolution of inventor networks in the Silicon Valley and Boston regions. Advances in Complex Systems, 10(01), 53-71.
18. Goyal, S., & Janssen, M. C. (1997). Non-exclusive conventions and social coordination. journal of economic theory, 77(1), 34-57.
19. Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 78(6), 1360-1380.

20. Hargadon, A. (2003). How breakthroughs happen: The surprising truth about how companies innovate. Harvard Business Press.
21. He, J., & Fallah, M. H. (2009). Is inventor network structure a predictor of cluster evolution?. Technological forecasting and social change, 76(1), 91-106.
22. Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. Advances in neural information processing systems, 14.
23. Kogut, B., & Walker, G. (2001). The small world of Germany and the durability of national networks. American sociological review, 317-335.
24. Leenders, R. T. A. (1996). Evolution of friendship and best friendship choices. Journal of Mathematical Sociology, 21(1-2), 133-148.
25. Liebeskind, J. P., Oliver, A. L., Zucker, L., & Brewer, M. (1996). Social networks, learning, and flexibility: Sourcing scientific knowledge in new biotechnology firms. Organization science, 7(4), 428-443.
26. McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
27. Milgram, S. (1967). Six degrees of separation. Psychology Today, 2, 60-64.
28. Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. The American economic review, 81(5), 1408-1418.
29. Nakao, K., & Romney, A. K. (1993). Longitudinal approach to subgroup formation: Re-analysis of Newcomb's fraternity data. Social Networks, 15(2), 109-131.
30. Powell, W. V. (1990). Neither Market Nor. Research in organizational behavior, 12, 295-336.
31. Rumsey-Wairepo, A. (2006). The association between co-authorship network structures and successful academic publishing among higher education scholars. Brigham Young University.
32. Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. Management science, 53(7), 1113-1126.
33. Snijders, T. A. (1998). Methodological issues in studying effects of networks in organizations. Computational & Mathematical Organization Theory, 4(2), 205-215.
34. Uzzi, B., Guimera, R., Spiro, J., & Amaral, L. (2006, August). Emergence: The dynamics of network formation. In annual meeting of the American Sociological Association, Montreal, Quebec, Canada.
35. Verspagen, B., & Duysters, G. (2003). In press. The small worlds of strategic technology alliances. Technovation.
36. Wang, M. (2009). Stimulative Policies for a Dynamic Network of Bioenergy Researchers: Identifying Strategic Collaborations [Doctoral dissertation, Texas Tech University].
37. Watts, D. J. (2004). Small worlds: the dynamics of networks between order and randomness (Vol. 36). Princeton university press.
38. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’networks. nature, 393(6684), 440-442.
39. Weesie, J., & Flap, H. D. (1990). Social networks through time. Isor.
40. Zeggelink, E. P., Stokman, F. N., & Van De Bunt, G. G. (1996). The emergence of groups in the evolution of friendship networks. Journal of Mathematical Sociology, 21(1-2), 29-55.

Appendix

Table 1: Poisson Regression

	Estimate
Intercept	-1.941*** (0.211)
Funding	0.236*** (0.025)
First degree	0.075*** (0.016)
Second degree	-0.002 (0.003)
Shared coauthors	0.447*** (0.078)
New authors	-0.092 (0.080)
Funding square	-0.006*** (0.001)
First degree x second degree	-0.001*** (0.000)

Table 2: Multinomial Logit Regression

	Group 1	Group 2
Intercept	-1.065*** (0.017)	-1.703*** (0.014)
Funding	0.021 (0.161)	0.103*** (0.054)
First degree	0.074** (0.031)	0.102*** (0.030)
Second degree	0.007 (0.007)	0.007 (0.005)
Shared coauthors	0.200 (0.128)	-0.486*** (0.054)
New authors	-0.242* (0.127)	0.492*** (0.055)
Funding square	-0.015 (0.014)	-0.003 (0.002)
First degree x second degree	-0.001*** (0.000)	-0.001*** (0.000)

Table 3: Logit Regression

	Estimate
Intercept	2.387*** (0.581)
Funding	-0.082*** (0.009)
First degree	-0.048*** (0.018)
Second degree	-0.017*** (0.008)
Shared coauthors	-0.966*** (0.188)
New authors	0.803*** (0.186)
Funding square	0.001 (0.006)
First degree x second degree	0.001 (0.001)

Figure 2: Network Evolution Under Fairness Rule

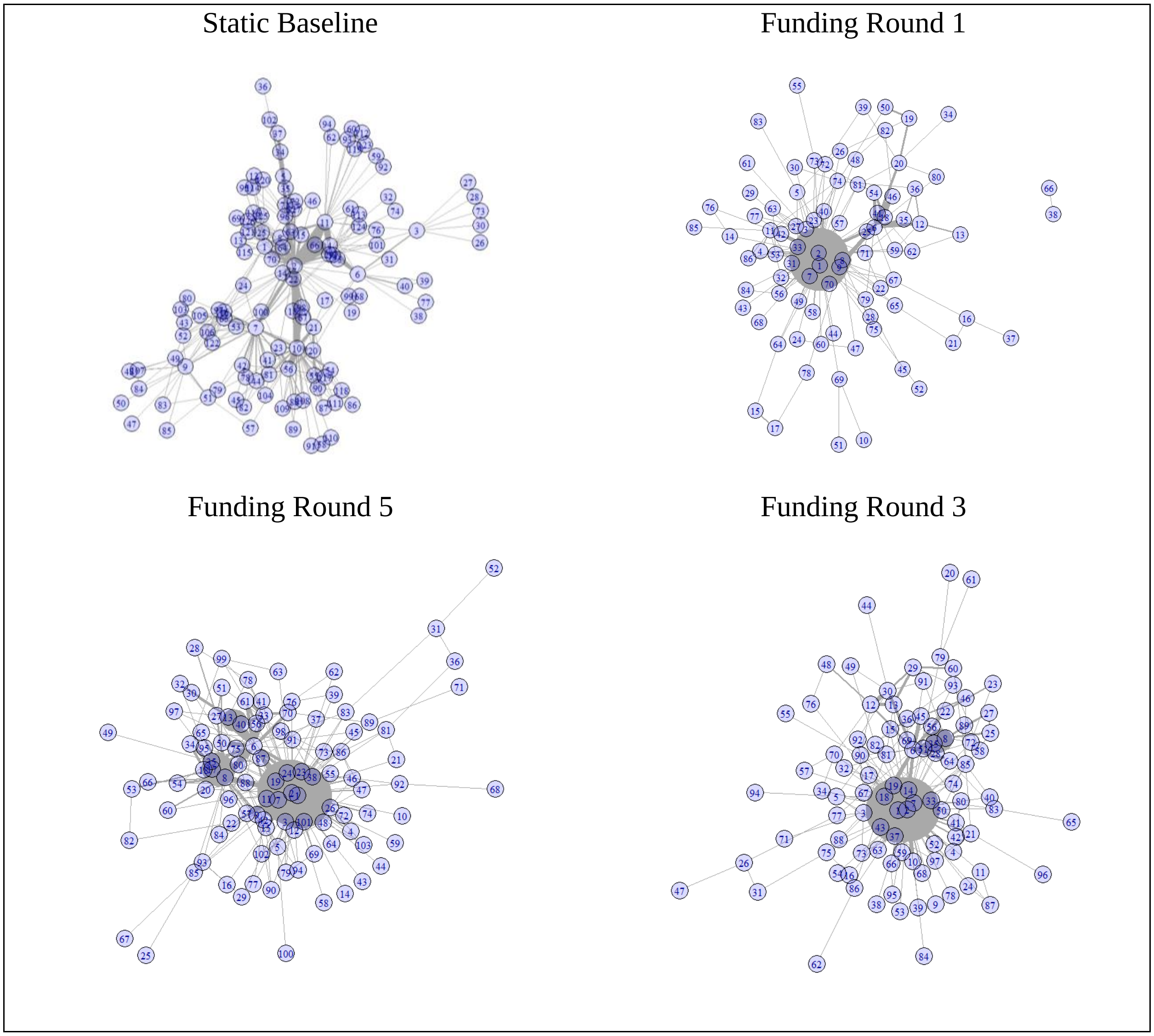


Figure 3: Distribution of Outcomes in Fairness Rule

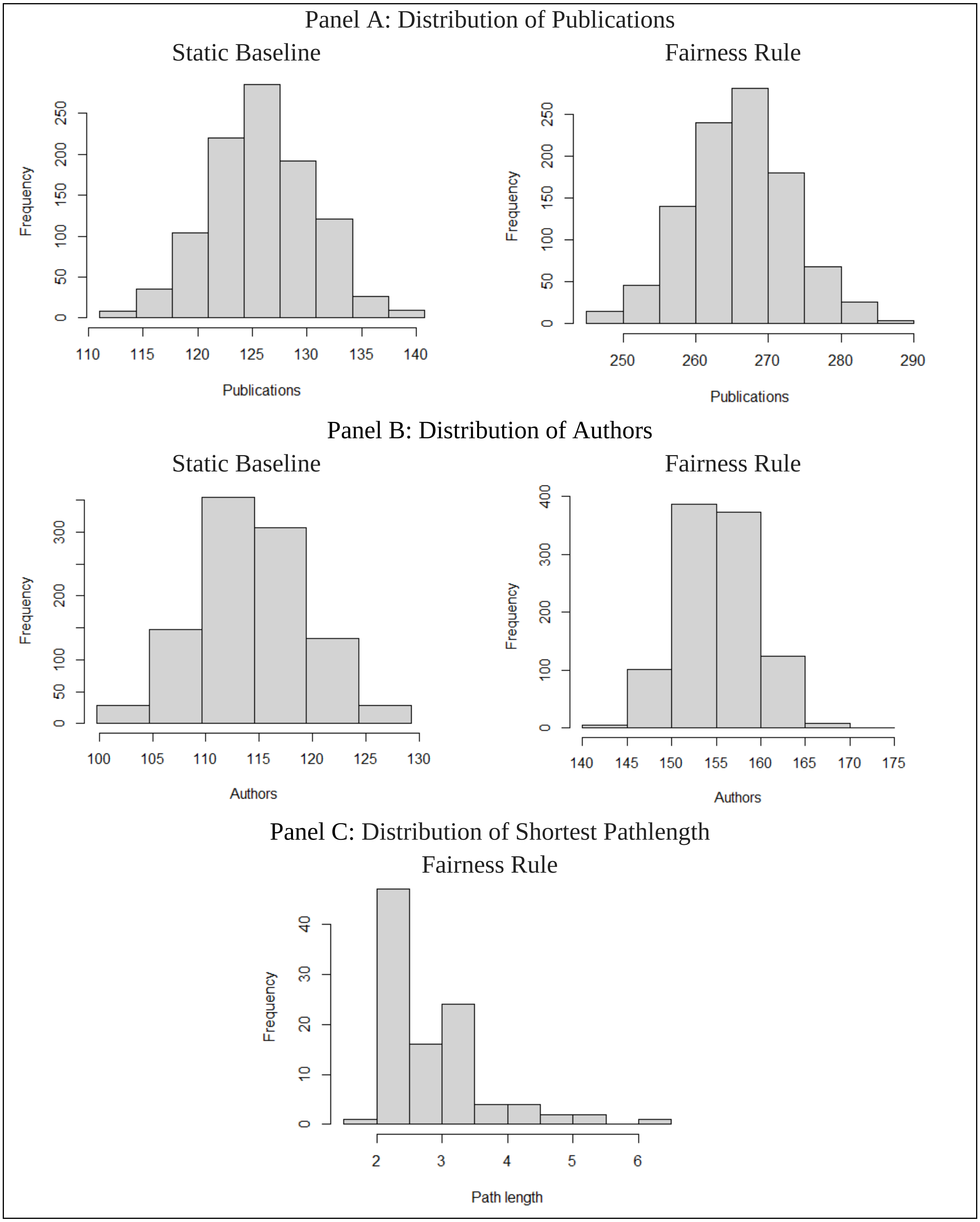


Figure 4: Network Evolution Under Smart Small World Rule

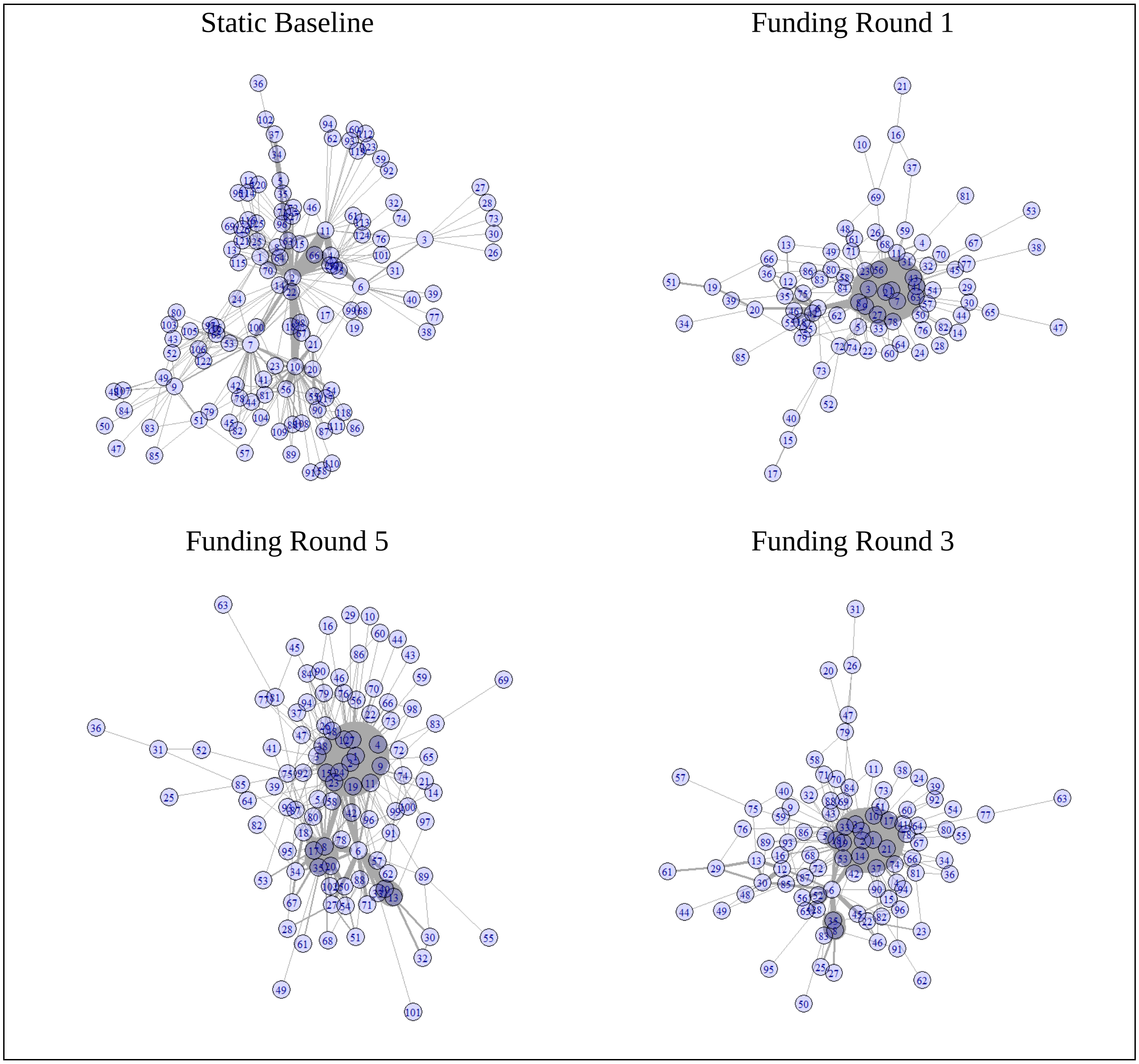


Figure 5: Distribution of Outcomes in Smart Small World Rule

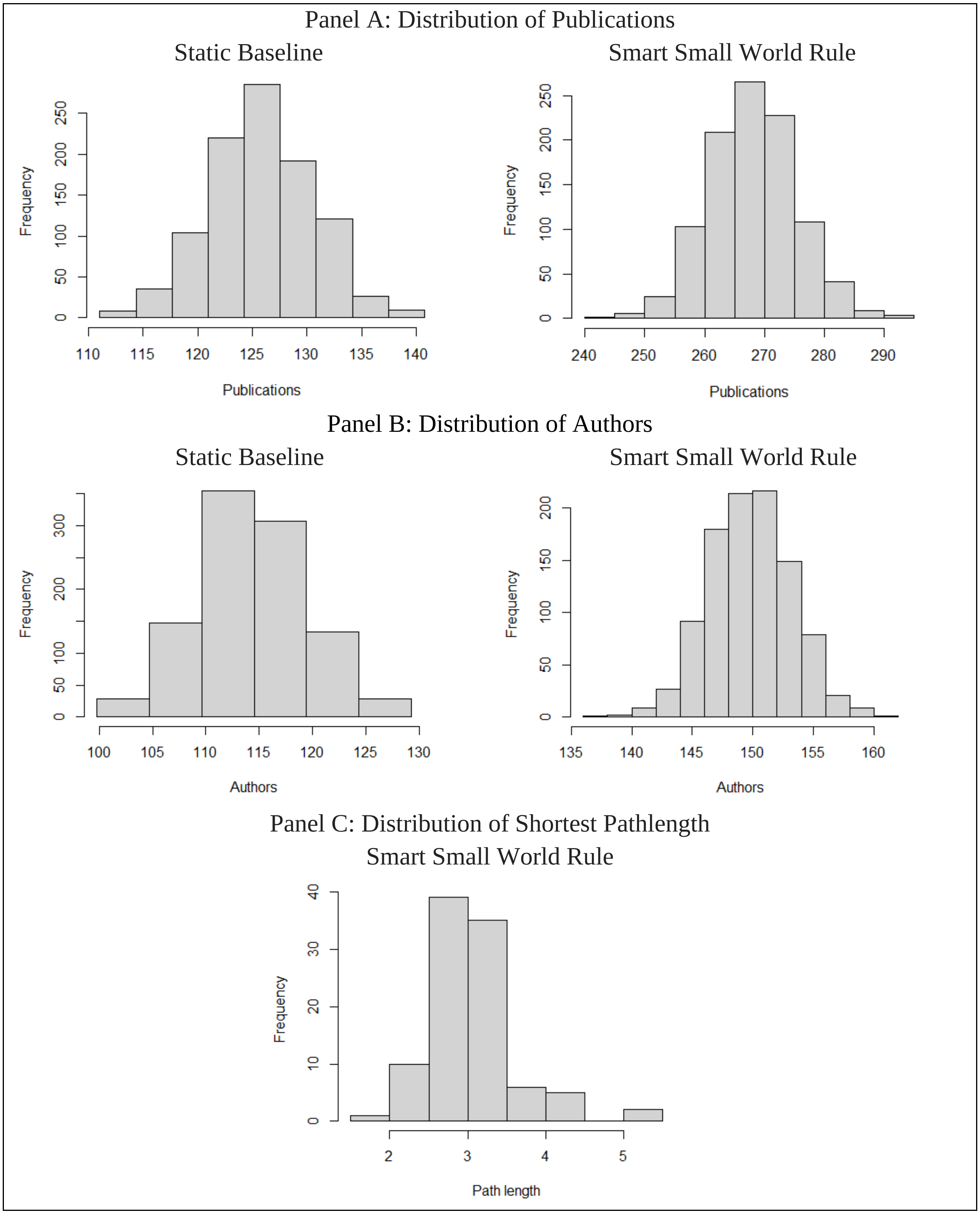


Figure 6: Network Evolution Under Direct Optimization

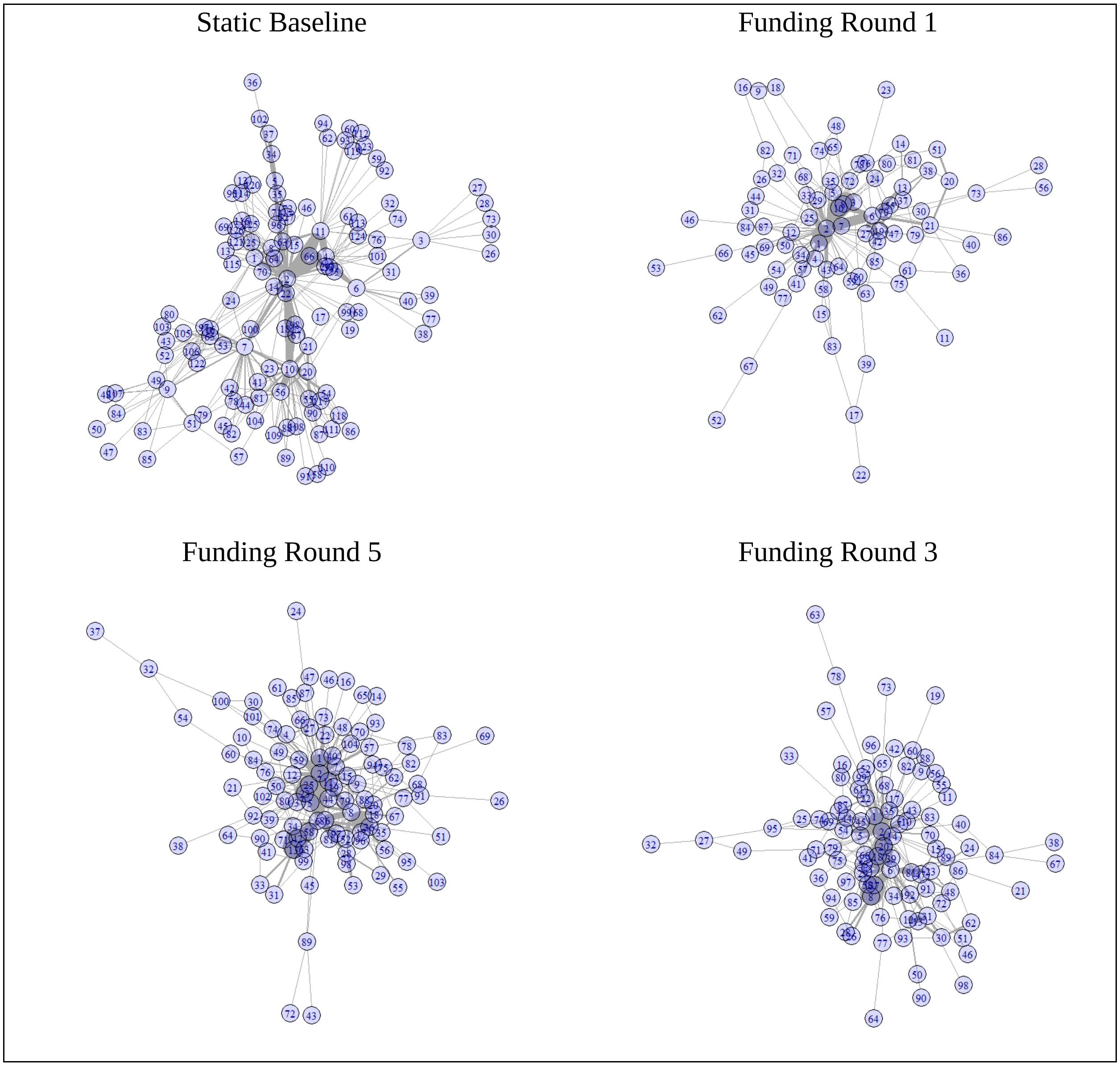


Figure 7: Distribution of Outcomes in Direct Optimization

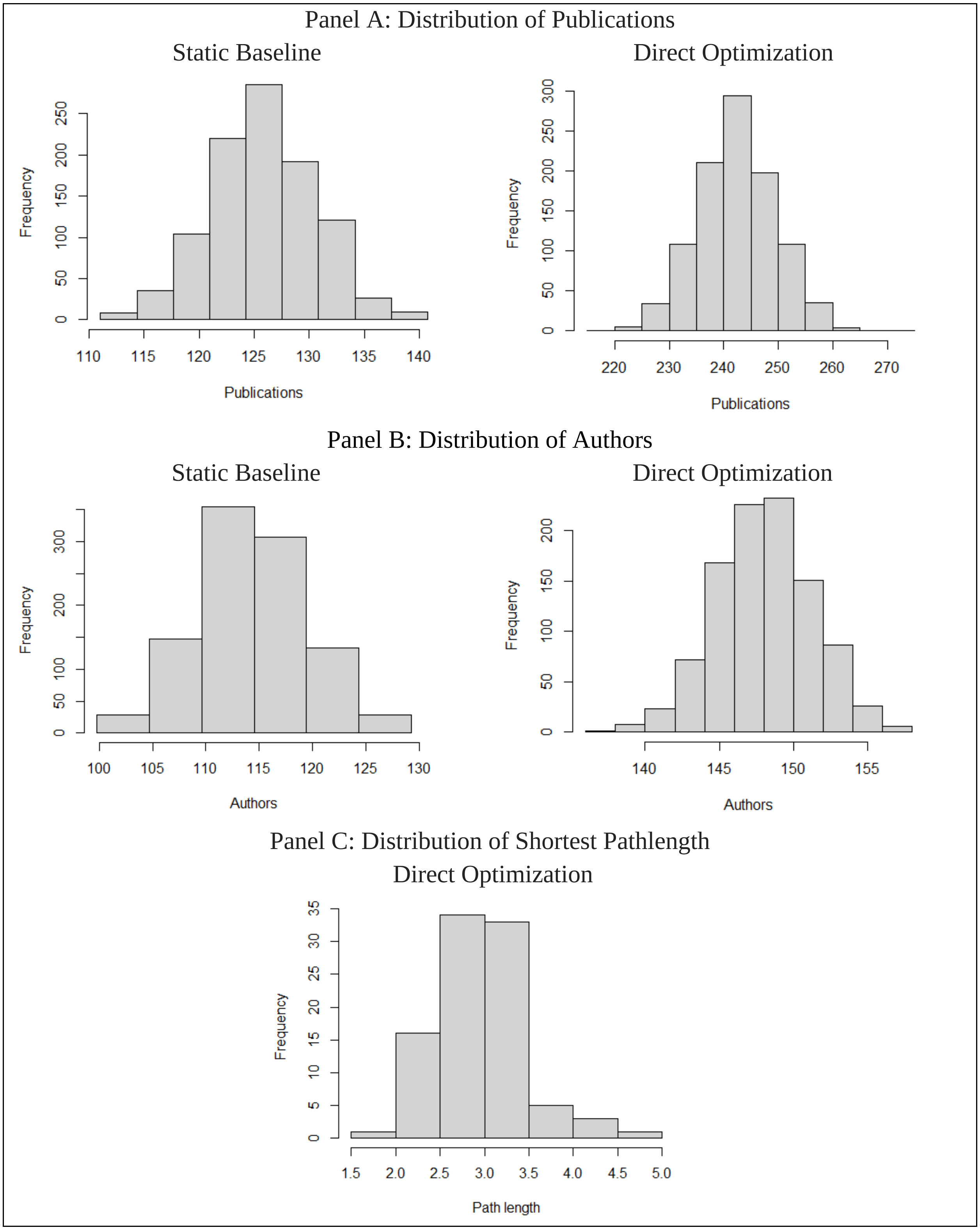


Figure 8: Network Evolution Under No Funding

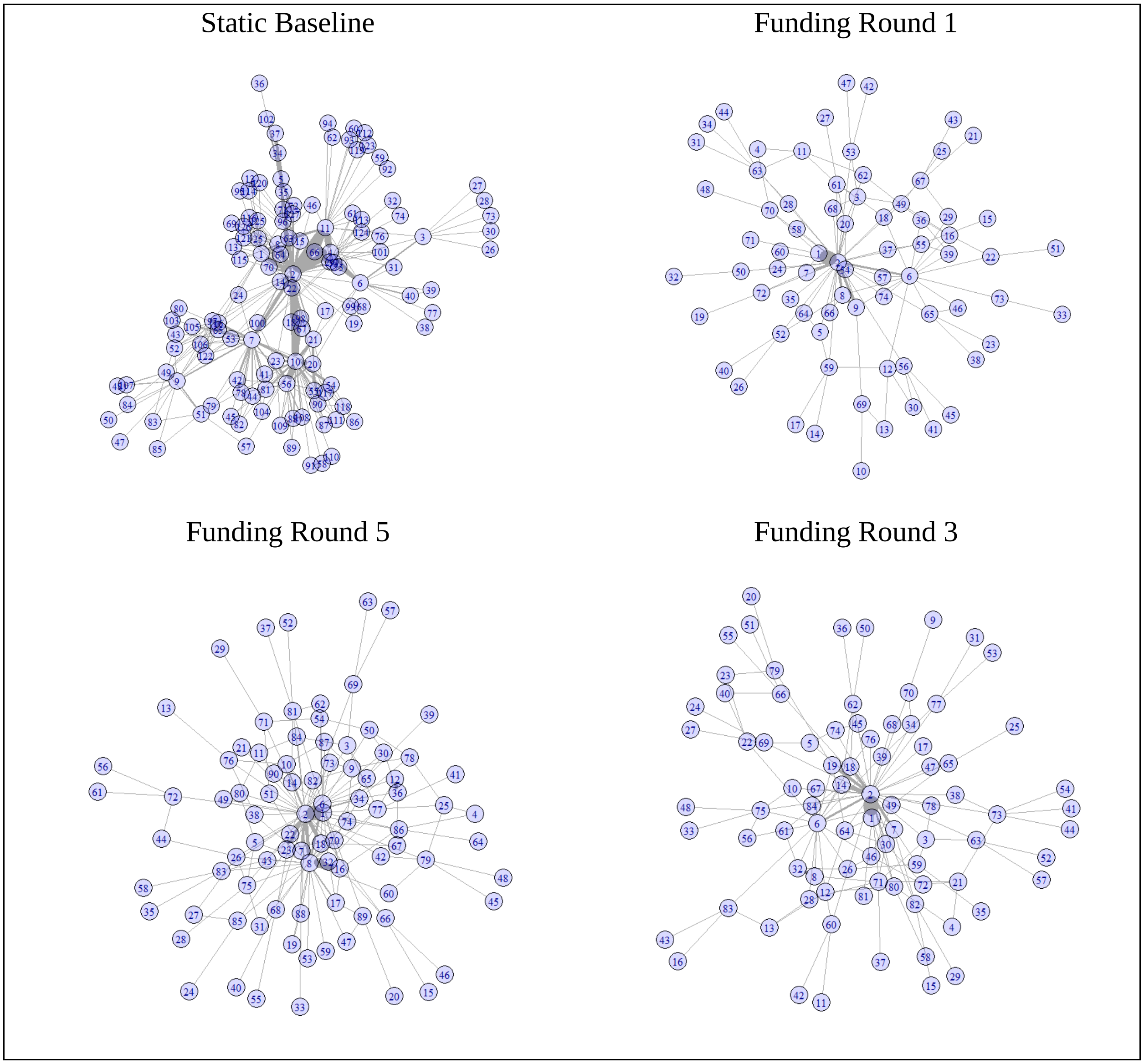


Figure 9: Distribution of Outcomes in No Funding