Who to Fund? Identifying Strategic Collaborations & Stimulative Policies for Dynamic Research Networks

Syed Fuad
Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX, USA
syed.m.fuad@ttu.edu
318D, Agricultural Sciences Building, Texas Tech University, 15th & Boston Lubbock, TX 79409, USA

Michael C. Farmer
Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX, USA

Who to Fund? Identifying Strategic Collaborations & Stimulative Policies for Dynamic Research Networks

Abstract

Black liquor is a by-product of papermaking. Currently, boiling black liquor in recovery boilers produces close to 0.56% of U.S. electricity supply; this potentially doubles if black liquor is gasified. Like many specialized bioenergy research communities, the Black Liquor Gasification research community is also small. Our data comprises the near universe of published work in this field, including information on all authors and funding that supported a given publication. We represent all collaborations as a social network to compare alternative funding strategies. One allocates funds to the most productive pairs. Others fund pairs that tighten overall network connectivity.

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. Direct Optimization funds coauthorship pairs expected to generate the highest expected number of publications. This increases expected publications by 92% and active researchers by 28%. Smart Small World funds pairs that create the largest number of researchers within one degree of separation. This strategy increases publications by 114% and adds 31% more researchers. Finally, a Fairness Rule minimizes overall pathlength, and increases publications by 111% and active researchers by 37%. Overall, this experiment suggests that research funding that strengthens overall research community connectivity generates the highest levels of research productivity and network connectedness.

Keywords: network dynamics, topology, black liquor gasification, small-world, research funding

JEL codes: C45, D85, O31

1. Introduction

Networks play a fundamental role in determining outcomes across various domains of social and economic interactions. Over the years, extensive field data has been employed through sharp mapping protocols to explain variation in information diffusion within and across networks. Such information could be something as mundane as a dinner invitation, or as substantial as information about job openings, the spread of disease or connectedness patterns in supply chains and energy distribution.

Given the ubiquity of networks, research on network topology has understandably relied on concepts borrowed from the natural and physical sciences. One such concept is the small-world property that explains, roughly speaking, how any two nodes in the network can reach each other through a very short sequence of acquaintances. Small-world networks are characterized by a distinct structure where most nodes are not directly connected but can be reached through a small number of intermediate steps. The core property that drives this phenomenon is the random long-range connection which allows nodes that are far apart in the network to be connected to each other, even if they are not directly connected. This is in contrast to regular lattice networks, where nodes are only connected to their immediate neighbors.

A similar, but distinct phenomenon is that of the 'strength of weak ties', the idea that weak ties or connections between individuals who are not closely connected or well-acquainted, can have significant social and informational advantages. This concept is particularly relevant in the context of social networks, where weak ties can bridge different social or professional circles and introduce new perspectives and opportunities. This happens because weak ties connect people who are part of different social circles, and often, the most useful resource is obtained from the occasional person whose connection is outside of homogenized, local associations.

The relationship between small-world networks and the strength of weak ties lies in the structural advantage that weak ties can provide within a small-world network. In a small-world network, weak ties between nodes contribute to short path lengths and efficient information flow. Weak ties can also act as the "shortcuts" that make the network more compact and facilitate rapid transmission of information. While distinct concepts, they both address the idea that networks with certain topological characteristics can enhance the flow of information, connections, and opportunities.

Numerous empirical studies speak to the consequential impacts of small-world networks across diverse, social networks. These include networks among corporate boards, strategic alliances, and academic collaborations (see Balconi et al., 2004; den Hamer and Frenken, 2021; Baruffaldi et al., 2016; Ayoubi et al., 2021). Social network researchers have suggested that networks among researchers in a given field enhance overall creativity, knowledge and innovation creation if that network exhibits a more generous small-world network structure (see Schilling and Phelps, 2007; Dahlander and McFarlan, 2013). More recently, the field of social networks has broadened to analyzing and understanding more nuanced dynamics of network evolution (Mannak et al., 2023). A variety of analytical tools have been developed, allowing researchers to analyze longitudinal relation datasets (Snijders and Doreian, 2010), stochastic actor oriented models for network dynamics (Snijders, 2017), coevolution, individual and joint node characteristics (Snijders et al., 2010; Hilbert et al., 2016). The common theme in 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 small-world network efficiently through weak ties.

Given the broad applicability and desirability of the small-world topology, we propose using strategic stimulation to directly install these properties into a network by targeting specific node collaborations rather than specific nodes. The dyadic approach is advantageous because it allows us to target specific pairs, who might not otherwise interact, based on some quantifiable parameter. Our approach is to construct rules about how given nodes or their connections mature and how they change based on the quality of their priors. Since individual outcomes are fraught with uncertainty, we find merit in providing an array of final expected outcomes for the entire network under a probability distribution obtained from hundreds of simulations that specifies individual variation in expected individual outcomes. Loosely, we take each node pair, its historical resource, and its position in the network to predict how that connection will change. We run hundreds of randomized simulations over the different estimated chances (drawing randomly from the probability distribution) that each specific outcome will succeed or fail and then map a distribution over network-wide patterns. We obtain a histogram of likely outcomes so that a policy maker can compare incidences of expected and desired outcomes.

This paper bridges several bodies of empirical literature - one being innovation and research and development; a second being the topological stylized networks literature; and a third about strategic searches in a network and the impact of topology. Within these fields, we address two specific research gaps. First, prior research has shown that research stimulation through funding generally has a positive impact on research network activity, interdisciplinary collaboration, and publication impact (see Porter et al., 2012; Beaudry and Allaoui, 2011; Zhao et al., 2018). However, there is limited understanding of how to *strategically target* stimulative funding to specific node pairs in a research network to achieve desirable individual- and network-level outcomes, such as research productivity and researcher connectedness. We also have limited insights into the consequences of node turnover (researcher entry or exit from network) on these outcomes (Chen et al., 2022). Second, while prior research has analyzed the relationship between network connectivity and innovation outcomes and has shown that specific network properties, such as the small-world property, are conducive to knowledge creation and diffusion, it is unclear whether stimulation through funding can be used as a policy vehicle to inject these desirable properties into a network.

To address these research gaps, we leverage the small-world topology to assess the impact of policy vehicles on improving research output in a social network of researchers in a specific bioenergy field. We use funding as a strategy to directly install small-world properties into a network by funding specific collaborations rather than specific researchers. We demonstrate the impact of node turnover as drivers of network evolution as they begin to exhibit small-world properties, which is another area of novelty of our study as most of the previous work on network evolution has focused on tie turnover (Zhang and Guler, 2020). Second, we provide an empirical study using traditional econometric approaches on the impact of funding injected strategically into a network to enhance its performance by facilitating specific network outcomes such as productivity, connectivity, and properties such as clustering, 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, coauthorship details, and the number of publications. This data is unique in its comprehensiveness and allows us to examine the impact of funding on the evolution of the research network in this field. Currently, black liquor, a by-product in the pulping process of wood into paper, is boiled. This energy conversion process produces close to 20 million MWh or 0.56% of the U.S. domestic electricity supply (U.S. Energy Information

Administration, 2019). 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 (Gebart, 2006).

Using publication information, we represent all pairwise scientific collaboration as connected 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. The probabilities are used to simulate the expected number of publications, total authors, average shortest path length, and clustering coefficient for the whole network in the first period given a level of funding. Based on the new network after first period, the same process is repeated four more times, each with 1,000 simulations, to obtain a distribution of the metrics for the network after five periods. We then obtain 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 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. The policy question we address here is the relative efficiency of different resources: direct funding as a resource, or collaboration capacity and overall connectivity. We compare their performance across various desirable outcomes against baseline metrics, including in distributions of productivity, total network nodes, their connectedness, average shortest path length, and clustering coefficient.

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, both desirable properties of small-world networks. 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 37%. An equivalent funding to the most prolific researchers (Direct Optimization) increases publication rates by 92% and researcher recruitment by 28%; and funding the most prolific collaborators (Smart Small World) increases publication rates by 114% and researcher recruitment by 31%. 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 highest levels of research productivity *and* network connectedness.

2. Background

2.1 Prior Literature

A substantial body of scholarly literature explores how economic outcomes are influenced by social network structures. Prominent examples include Boorman (1975) and Montgomery (1991), who found a relationship between labor market outcomes and social network; Ellison and Fudenberg (1995) found that the structure of communication can affect a consumer's purchasing decisions; Cross and Parker (2004) showed that organization of workers influences a firm's efficiency; and in evolutionary game theory, Veloz et al. (2014) and Scata et al. (2016) demonstrated the influence of network structure on possible coordination among agents.

Network evolution and co-evolution, topology, and its relationship to diffusion and knowledge resources have been topics in the field for quite some time. As scientists seek to construct models of social processes 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. Such dynamic 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 importance of such problems has prompted a good deal of methodological research on network variation. Examples of this research are common and extensively cited: Powell et al. (1996), and Powell et al. (2005) are obvious examples. More recent (and hence less frequently cited) examples include Ayoubi et al. (2019), Baruffaldi et al. (2016), Ayoubi et al. (2021), Kim et al. (2022) and Rawlings et al. (2015). Network variation has been tested in the context of traditional network analysis (Azoulay et al. (2017) and more innovative relational event models (Lerner and Lomi, 2020).

2.2 The Notion of Small-World and The Strength of Weak Ties

The small-world notion posits that any two people in the world who are randomly selected are connected to each other by a small number of intermediate links. This phenomenon was first observed by Milgram (1967), who conducted a quantitative survey in which 296 individuals in Nebraska were asked to deliver a letter to a specific person in Boston whom they did not know. The study concluded that each pair of people in the world is separated, on average, by six intermediate acquaintances. This has since been formalized by Watts and Strogatz (1998) who studied 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 – these properties have since come to characterize small-world networks.

Prior research has also established how small-world networks foster innovation in general (den Hamer and Frenken, 2021, Schilling and Phelps, 2007, Uzzi and Spiro, 2005). Many empirical studies have located the small-world property operating in various fields of social networks, including in corporate boards, industries, strategic alliances, investment bank syndicates, email lists and mentoring networks, and scientific collaboration networks (see Balconi et al. 2004, Feeney and Bozeman, 2008; Goyal et al., 2006, Schnettler, 2009, Prell and Lo, 2016, and Anjos and Reagans, 2013).

2.3 Why Connections Matter?

The importance of connections in determining outcomes of a network in a dynamic process begins with the idea that social networks link nodes to one another (Mannak et al., 2023). At a granular level, connections provide information and deliver resources across network systems to their final users. On a collective scale, as individual connections act and operate in sequence over a network, connectivity can change with nodes entering and exiting a network over time, which can depend on the ability of a connection positioned in a network to acquire resources for its nodes. Within a model, independent nodes are flexible and can interact, thus facilitating observations on the relationship between the behavior of each agent and the status of the resource.

This paper describes the development of a network with resource management implications. An intelligent agent-based simulation of individual action and overall group performance in a simplified *commons rules* is presented. Our objective is to support policymakers in identifying strategic connections that will facilitate greater research output and network connectivity. We hypothesize that individual node pairs 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 applied to certain types of networks, while still respecting the basic small-world insight of long-range ties. We argue that social network mappings, which are becoming increasingly sophisticated with the development of advanced data collection and mapping tools, contain 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 fully utilize this data, especially individual's information regarding how they acquire resources and accomplish specific tasks.

2.4 Why Black Liquor Gasification (BLG)?

Black liquor, a by-product in the pulping process of wood into paper, contains 15–17% solids, a high organic content consisting of dissolved organics from the wood and spent pulping chemicals (Bajpai, 2016). It has an approximate energy content of 14 MJ/kg dry solids, which is about half of the energy content of 1 kg of coal equivalent (Gebart, 2006). It is desirable to recover and recycle these chemicals for both environmental and economic reasons. The current black liquor handling technology uses a Tomlinson recovery boiler to recover the energy and chemicals from black liquor. However, the efficiency of the current recovery technology is relatively low because of black liquor's relatively high water content, which limits combustion efficiency (Bajpai, 2016).

Black liquor gasification (BLG) is an alternative recovery technology developed in the 1960s. Compared with conventional recovery technology, BLG has the potential to extract more energy content from black liquor and increase the amount of energy generated at the pulp mills and sold to the power grid by two to three times (Gebart, 2006); this a substantial bioenergy contribution from this one research activity involving relatively few researchers. However, the investment for a full-scale pressurized BLG process unit is 60-90% times higher than for a new conventional recovery boiler (Bajpai, 2008). While current BLG technology increases pulping process energy recovery by about 10%, this efficiency has to be improved further to advance the commercialization of BLG in the pulp and paper industry.

Gasification technology research was primarily funded by the U.S. Department of Energy until the early 2010s. A few gasification technologies have been tested on pilot scale but most of them were abandoned due to technical inferiority and very few are now at commercial stage as funding has mostly dried out. (See Naqvi et al., 2010 for a comprehensive review of the research and development of BLG technologies.)

Our choice of this field of research is guided by two reasons. First, the small network of researchers involved in this field has the benefit of allowing us to collect an almost exhaustive list of researchers, their publications, authorship, funding, funding sources, etc. which ultimately allows us to represent all collaborations as a social network. Second, the recovery boilers, most of which were established in the 1970s and 1980s, are near to complete their functional life spans, are becoming energy and environmentally inefficient and need to be replaced in the next few decades, preferably with gasifiers (Naqvi et al., 2010). Thus, continued funding is needed in this field to promote research and development to solve critical issues related to BLG technology for commercialization of biorefinery systems at pulp mills.

3. Data Description and Methods

The data in this study is obtained from a survey of the 20 top researchers in black liquor gasification, including the top nine (9) most published researchers in this field who are collectively responsible for 38% of all publications in this topic area. The initial list of researchers is obtained from *"Vantage Point"*, a commercial program developed at the Georgia Institute of Technology in the 1990s to track publication record of individual authors in high tech disciplinary areas. The initial list was subsequently verified through reconciliation checks with domain experts to confirm that it is indeed comprehensive and covers scientists publishing the bulk of the research in this field (Sinquefield, 2008; Frederick, 2008). Each author in the list was then contacted, and based on the information provided, we extracted information on the list of publications by each researcher, date of publication, the names of all coauthors, second degree coauthors (coauthors of coauthors), the presence of any new authors entering the network through a particular paper, and the funding level of the corresponding project and papers resulting from each funded project. The final survey data includes information on a total of 126 peer-reviewed publications, published in 40 journals by 127 researchers from 60 institutions.

3.1 Independent Variables

We represent every author as a node in a network; if any two authors have at least one shared publication, we connect the pair as a collaboration. For every collaboration, we construct the following independent variables using the original data:

1. Independent Variable – *Funding*: total funding level for any two authors' publications. Suppose author A publishes 15 papers and author B publishes 10 papers, and 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 the variable will take a value of $(m+n)/20$, defined as the funded dollars (in $100,000) per publication.
2. Independent Variable – *First degree*: this is the total first degree coauthor relations in the network for any two authors. If 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 (5+6) 11. If there are duplications, only unique coauthors are counted.
3. Independent Variable – *Second degree*: this is defined as the total second degree relations, or those two degrees of separation from an author-pair. We define second degree relations as follows: if author A is only indirectly connected to 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. In other words, the total number of persons two degrees away count as the number of unique 'coauthors of coauthors'.
4. Independent Variable – *Shared coauthors*: this is the number of total coauthors from any two authors' shared publications. Suppose authors A and B publish 5 papers together, and these 5 papers involve 15 unique coauthors, then the value of this variable is equal to 15. This is therefore the number of other coauthors a particular author-pair shares directly in their joint work.
5. Independent Variable – *New authors*: this is the 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 publish 10 papers together, and there are 4 new authors in these 10 papers, then the value of this variable is equal to 4).
6. Independent Variable – *Funding square*: this is the square of the *Funding* variable.
7. Independent Variable – *First degree x second degree*: this is the interaction between the independent variables *First degree* and *Second degree* coauthors.

3.2 Dependent Variables and Regression Models

A key element in our research is the movement of researchers into and out of the network as drivers of network dynamics. We refer to this movement as network evolution. We use the following regression models (depending on the nature of the dependent variable) to guide network productivity and evolution of connectivity within the network:

- A Poisson regression is used to estimate the productivity for a connection. The dependent variable is the number of shared publications by an author pair (e.g., if authors A and B publish 5 papers together, then the dependent variable, y, is equal to 5), and the unit of analysis is every possible pair of researchers. With the probability from the Poisson regression, we can first simulate the productivity of a connection, and then 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. We have three possible values for the dependent variable: (1) where the coauthor pair does not recruit any new researcher in their shared publications; (2) where the pair recruits a new researcher who publishes only once; and (3) where the pair recruits a new researcher who publishes more than once. With the probability of entry for each author pair, we can then simulate the number of new authors to enter the network.
- A Logit regression is used to estimate the number of researchers who permanently exit a network. The dependent variable is (1) if the researcher pair breaks and (0) if not. First, we find the probability of whether an author pair connection is broken or not. Second, if this connection is broken, we find which author in the pair exits the system. Then with the probability of exit for each pair of authors, we can simulate how many authors will exit from the whole network.

The summary statistics for the three regression models are presented in Table 1. Additional details about how the data for the models is obtained from the raw data of publication information is presented in section 1 of *Appendix A: Supplementary Material.*

4. Results

4.1 Parameter Estimation

4.1.1 Productivity Model

The first model for observing the productivity of a network is estimated by Poisson regression. The unit of analysis is every author connection or pair of researchers, the dependent variable is the discrete number of publications by the pair, and the independent variables are the seven variables defined in Section 4.1. The method also provides a discrete probability of publishing a given number of papers, which can subsequently be employed in randomized simulation.

Table 1: Summary Statistics

	Poisson Model	**Multinomial Logit**	**Logit Model**
Dependent variable	0.426 (0.566)	2.170 (0.838)	0.478 (0.499)
Funding	0.085 (0.063)	0.446 (0.413)	0.446 (0.413)
First degree	11.501 (8.448)	26.740 (17.571)	26.740 (17.571)
Second degree	59.832 (20.389)	76.450 (26.718)	76.450 (26.718)
Shared coauthors	21.435 (106.051)	53.695 (29.714)	53.695 (29.714)
New authors	10.167 (50.757)	50.689 (29.082)	50.689 (29.082)
Funding square	1.121 (1.586)	3.692 (9.246)	3.692 (9.246)
First degree x second degree	775.411 (785.536)	2,424 (2,090)	2,424 (2,090)
N	*8,001*	*682*	*682*

The results are presented in Table 2 and suggest the following. In most cases, the independent variables are statistically significant with a p-value less than 0.05, except variable for second degree. We are primarily interested in the effects of funding, and less so in the first and second degree relationships, new coauthors, and shared coauthors; although as the table reveals, most other variables are also significantly related to the dependent variable.

The effect of funding level on publications is significantly different from zero; the result can be interpreted to mean 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 increases by one unit, measured in $100,000. In other words, the probability that any two authors will successfully coauthor a paper increases by a factor of 1.266 or 26.6% as the research funding to the authors increases by $100,000.

First degree relationship is also an important variable associated with the publication probability. The probability that pairs of researchers will publish increases by 7.79% for every additional unit of first degree relations. By the same criterion, the variable of new authors increases the probability of success. For every unit increase in the number of new coauthors a pair of researchers brings, the probability of publication success improves by 9.64%.

Table 2: Models to Estimate Productivity and Connectivity Outcomes

	Poisson Model	**Multinomial Logit Model**		**Logit Model**
	Outcome: Productivity	**Outcome: Entry**		**Outcome: Exit**
		$\boldsymbol{y = 1}$	$\boldsymbol{y = 2}$	
Intercept	-1.941*** (0.211)	-1.065*** (0.017)	-1.703*** (0.014)	2.387*** (0.581)
Funding	0.236*** (0.025)	-0.021 (0.161)	0.103*** (0.054)	-0.082*** (0.009)
First degree	0.075*** (0.016)	-0.074** (0.031)	0.102*** (0.030)	-0.048*** (0.018)
Second degree	-0.002 (0.003)	0.007 (0.007)	0.007 (0.005)	-0.017*** (0.008)
Shared coauthors	0.447*** (0.078)	0.200 (0.128)	-0.486*** (0.054)	-0.966*** (0.188)
New authors	0.092** (0.040)	-0.242* (0.127)	0.492*** (0.055)	0.803 (0.686)
Funding square	-0.006*** (0.001)	-0.015 (0.014)	-0.003 (0.002)	0.001 (0.006)
First degree x second degree	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	0.001 (0.001)
N	*8,001*	*682*		*682*
Pseudo R-squared	*0.3879*	*0.0851*		*0.3559*
Log-likelihood	*-1271.0526*	*-1167.8766*		*-289.1522*
LR Chi-squared	*3404.36****	*225.54****		*179.98****

*Note: Standard errors are in parentheses. *p < 0.1, **p < 0.05, ***p < 0.01.*

What is striking in the result is the power of additional coauthorship, where an additional coauthor increases publication success by 56.36%. 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 coauthorship history might be a source for a viable stand-alone policy.

4.1.2 Entry Model

The 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 pair of researchers, except that only authors who share a publication are paired (unlike the previous model, where all possible combinations of authors are paired).

The multinomial regression gives the parameters estimates for entry model, which is used to get the probability that any pair of researchers could recruit a new researcher into the system. Here, group 1 indicates coauthor pairs who do not recruit additional authors ($y = 1$) and group 2 indicates coauthor pairs who recruit additional authors who only publish once ($y = 2$). Because we only care about how many new researchers enter the system permanently, we focus on analyzing the probability of group 3, which is also the reference group.

The results are presented in Table 2. The exponentiated coefficients (relative risk ratio) for funding, first degree, and new authors are less than 1 for group 1, indicating that as the value of these variables increase, the reference outcome (that the author pair recruits researchers permanently) is more likely.

4.1.3 Exit Model

The third model is used to estimate the probability of exit for each pair of researchers using a logit regression. The dependent variable is either zero, meaning the connection will not break, or one meaning that the connection 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, we then estimate 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 2. Most independent variables are statistically significant at the 5% level and have a negative sign, implying that the exponentiated coefficients would be less than 1 and so increasing their values would result in decreasing the odds of the connection breaking. In sum, 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.

4.2 Static Baseline Simulation

With the regression results, we now simulate the baseline network to better illustrate the simulation rules and principles. (A more detailed description of the process is presented in section 2 of *Appendix A: Supplementary Material*.)

4.2.1 Poisson Simulation

We first calculate the Poisson probability $\left(P(x = h)\right)$ for each possible number of publications $(h = 0, 1, 2, \dots, 19, 20)$ for each author connection in the data. 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 in the data, and then take the sum of all simulated numbers and obtain a total number of connections of 688 for the whole network.

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. To avoid this duplication problem, we 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 scenario. This process is repeated 1,000 times to eventually obtain a distribution of total publications.

4.2.2 Multinomial Logit Simulation

Next, we simulate the number of new author entries 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 does not recruit a new coauthor, and vice versa. However, the same duplication problem exists, and can be similarly adjusted to estimate 25 new researchers entering the network.

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

This is repeated 1,000 times to obtain a distribution of the number of new author recruitment.

4.2.3 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 probability. 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 that 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. Like the previous cases, this process is also repeated 1,000 times to obtain an average estimate of 38 authors leaving the network.

In sum, an average total of 25 new authors enter the network and 38 authors leave, so the Static Baseline network comprises 114 authors (13 fewer than the original network) publishing 126 papers. This network is illustrated in Panel A of Figure 1 and forms the foundational network from which other networks evolve.

4.3 Stimulative Dynamic Funding Policies

We evaluate three alternative funding policies: Fairness Rule, Smart Small World Rule, and Direct Optimization. These initial plausible policy choice rules differ by which specific pairs of nodes are injected with funds and 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 the different funding scenarios.

4.3.1 Fairness Rule

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

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

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 author connections that display small pathlengths, so in reality, the policymaker is not limited to a small set of options.

In our case, we choose 5 author connections with the minimum shortest pathlength and give each pair $100,000 and iterate the simulation five (5) times. This is relatively trivial to do, since our data are already organized by author pairs as units of observation. A five period iteration implies $500,000 of additional capital for each pair, with five pairs receiving a total of $2,500,000 in five rounds of funding. This is equivalent to 22% of the total funding in the raw data. Panel B in Figure 1 demonstrates the final network after five funding rounds. Each node represents an author, the vertex (edge) connecting any two nodes indicates whether the nodes published together, and the thickness of the vertex indicates the strength of a relation (the more shared publications, the stronger the relation).

Table 3 presents the mean and standard deviation of the network outcomes and properties after five simulation rounds. We observe a significant increase in publications after funding the most efficient collaborators; the mean number of publications increases by 111% to 266 (8.17 s.d.) publications compared to the Static Baseline case of 126 (4.81 s.d.) publications. This is a significant increase in productivity in response to only 22% more funding over five cycles. The mean number of authors in Static Baseline is 114 (5.49 s.d.), while with additional stimulative funding increases this mean to 156 (4.24 s.d.), equivalent to a 37% increase. The standard deviations reported in parenthesis are relatively low, suggesting the network outcomes are quite stable and resilient.

The third measure we evaluate is the average shortest pathlength, also called geodesic distance. This is one of the most robust measures of network topology and the most common way of ascertaining the small-world nature of network (Humphries and Gurney, 2008). The shortest pathlength is defined as the path with the minimum number of vertices 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 network is characterized by a small shortest pathlength, which facilitates efficient information transfer within a network (Watts and Strogatz, 1988). In this case, the mean of shortest pathlength reduces significantly from 3.898 (0.308 s.d.) in the Static Baseline to 2.897 (0.796 s.d.) in response to funding, which suggests that following this policy may stimulate authors to build more first degree connections in addition to publishing more.

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 clustering coefficient quantifies the abundance of connected triangles in a network.[1] The small-world index is a calculated variable measured by dividing the clustering coefficient by the shortest pathlength (Eslami et al., 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. Under the Fairness Rule funding policy, the clustering coefficient and density are higher than the Static Baseline case.

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. Leveraging this strategy generates 111% more publications and recruits 37% more researchers to the network with only 22% more funding. The results also suggest that it is possible to inject funds strategically

[1] The clustering coefficient is formally defined as: $C = \frac{1}{n}\sum_{i=1}^{n} C_i$, where n is the number of nodes in a network and $C_i = \frac{number\ of\ triangles\ connected\ to\ node\ i}{number\ of\ triples\ centered\ around\ node\ i}$, where a triple centered around node i is a set of two vertices connected to node i.

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.

4.3.2 Smart Small World Rule

The second stimulative funding policy is the 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 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}:

$$Max \sum C_{i,k} + \sum C_{j,m} + C_{ij} \tag{2}$$

Simply put, 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 linked is as large as possible. Like the last policy, we choose 5 pairs of researchers with the highest first degree connections in the network. Each of the pairs is provided with $100,000, and the simulation process is repeated five times. Panel C in Figure 1 illustrates the final network in response to this funding policy.

From Table 3, we see a significant increase in the number of publications after repeating this funding over five cycles; the mean of the number of publications for Smart Small World Rule increases by 114% to 270 (7.11 s.d.) publications compared to the Static Baseline case of 126 publications. The mean of total authors in Static Baseline simulation is 114, while providing additional funding increases this to 149 (3.43 s.d.), equivalent to a 31% increase.

The mean of shortest pathlength also reduces from 3.898 to 3.073 (0.589 s.d.), which suggests that funding the most prolific collaborators increases publications, author connectivity and makes the research network more compact and efficient, all properties of a small-world network.

4.3.3 Direct Optimization

The third policy criterion is the Direct Optimization Rule. If the policymakers' 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 a single connection. It uses information beyond connectivity, based on expected collaboration success. In this case, collaboration success means the number of 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 vertex 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 $P(S_{ij}) = S(C_{ij}(x))$ 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):

$$Max \sum P(S_{ik}) + \sum P(S_{jm}) + P(S_{ij}) \quad (3)$$

The Direct Optimization Rule differs from the Smart Small World rule, which maximizes the number of connections between the most highly connected nodes, by 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 choose 5 pairs of researchers who demonstrate the highest expected publications in the network and assign $100,000 to each pair and iterate the same process over five funding periods. Panel D of Figure 1 illustrates the final network under this funding policy of funding the most productive researcher pairs.

The results in Table 3 show a significant increase in publications after funding author pairs with the highest expected publications. The mean number of publications for Direct Optimization increases from 126 to 242 (5.93 s.d.), 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 Fairness Rule and Smart Small World Rule, this policy resulted in a relatively smaller 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 28% to 146 (3.01 s.d.), 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 the 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 (0.952 s.d.).

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.

Table 3: Summary Diagnostics

	Avg. # of publications	Avg. # of authors	Avg. shortest pathlength	Clustering coefficient	Small world index
Static Baseline	126 (4.81)	114 (5.49)	3.898 (0.308)	0.442 (0.022)	0.113 (0.011)
Fairness Rule	266 (8.17)	156 (4.24)	2.897 (0.796)	0.499 (0.020)	0.172 (0.025)
Smart Small World Rule	270 (7.11)	149 (3.43)	3.073 (0.589)	0.500 (0.029)	0.163 (0.016)
Direct Optimization	242 (5.93)	146 (3.01)	2.952 (0.545)	0.480 (0.014)	0.163 (0.019)
No Funding	116 (20.28)	90 (18.26)	2.992 (0.674)	0.357 (0.121)	0.119 (0.052)

Note: Standard deviations in parenthesis.

4.3.4 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 Panel E of Figure 1; we observe a marginally smaller number of publication outputs (126 to 116; an 8% decrease), significantly smaller number of authors (114 to 90; a 21% decrease), and interestingly, a reduction in shortest pathlength (from 3.898 to 2.992).

5. Policy Summary and Discussion of Results

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.

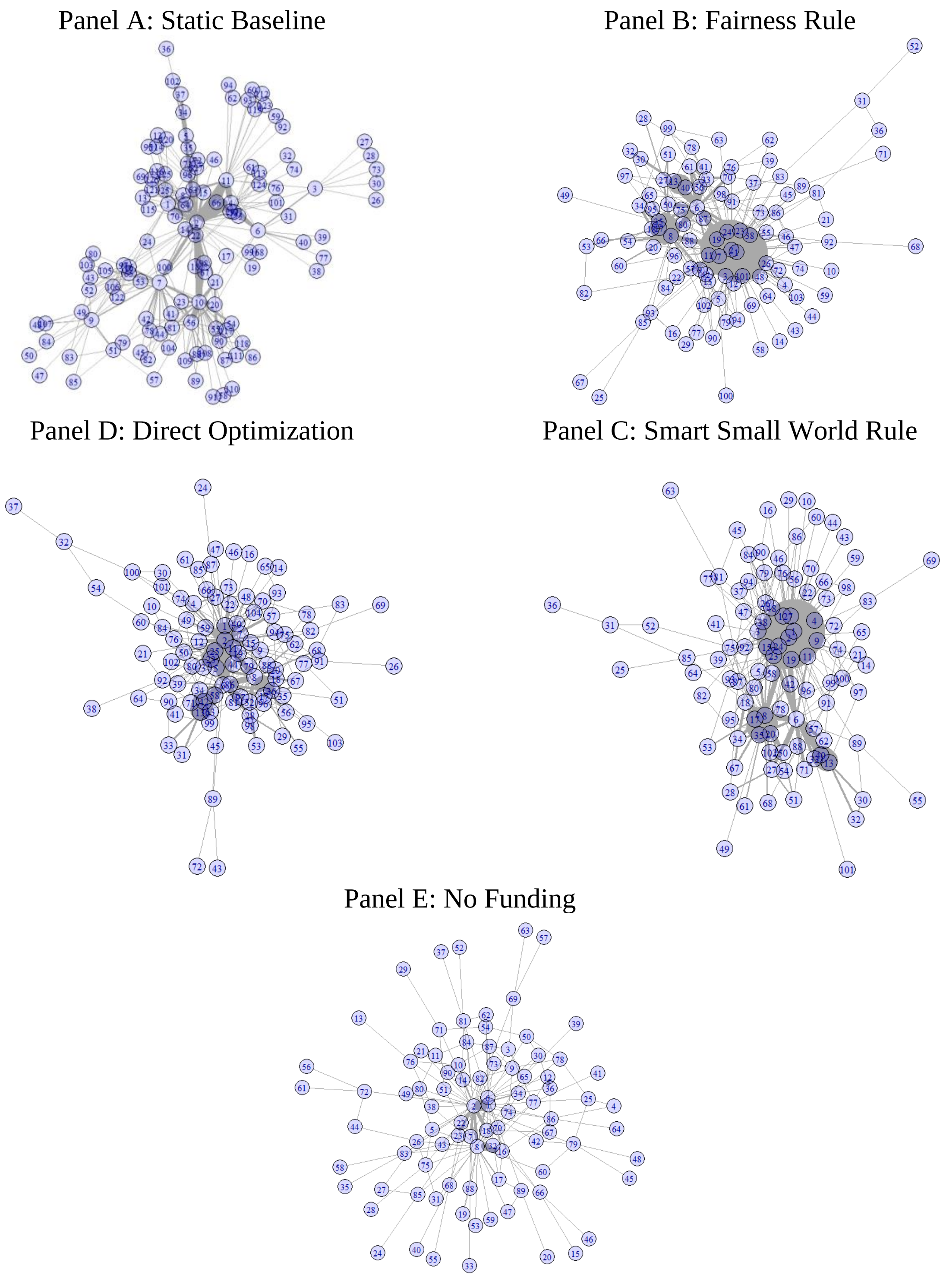
The Smart Small World Rule funds author pairs with the highest number of first degree connections and yields the highest number of expected publications. Direct Optimization funds author pairs with the highest number of expected publications and produces the smallest increase in the number of expected publications and author recruitment equivalent to the level of Smart Small World Rule.

All funding policies yield shorter average pathlength and higher clustering coefficient compared to the Static Baseline scenario, which itself exhibited nuanced small-world properties including the organization of clusters; the targeted funding policies then substantially reinforced and enhanced the original small-world properties. In all funding policies, as more authors enter the network, they are initially connected only to the coauthors by whom they are recruited. As the funding cycle progresses, new authors begin to form connections with other authors outside those who initially recruited them, resulting in the growth of multiple interconnected clusters (Figure 1). This predictably decreases the length of connectivity for new authors to others in the network and appears to be a case of the strength of weak ties phenomenon. These outcomes emerge as the networks have a high density of local connections and a small number of long-range connections which connect researchers from different clusters.

Interestingly, we observe substantial differences in the outcomes from alternative funding strategies, especially between strategies that optimize publications and those that optimize connectedness. For instance, the Fairness Rule and Smart Small World Rule, both of which optimize network connectedness, generate average publications of 266 and 270 respectively; the Direct Optimization method, which funds the most productive author pairs, generates an average of 242 publications. The standard deviation of the Direct Optimization is 5.93, implying the average publication outcome for this method is over 4 standard deviations away from that of the Fairness Rule, and almost 5 standard deviations from the Smart Small World Rule. In terms of the number of authors in the network, the outcome of the Direct Optimization is 1 standard deviation from that of the Smart Small World Rule and over 3 standard deviations from that of the Fairness Rule.

Two prior studies on similar topics (see Prell and Lo, 2016 and Robins et al., 2005) compare different strategies for generating topological structures using knowledge resources, and both make a strong case for anything within 2 standard deviations being insignificant. By that benchmark, our simulations reveal that strategies that are different in terms of which node pairs they target for funding yield significantly different results in terms of network productivity, connectedness, and small-world properties.

Figure 1: Network Evolution Under Alternative Funding Rules



Note: Figures show the Static Baseline network, and the empirical networks after the final funding round. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

One important thing to note is that the networks under different funding scenarios cannot always be reliably compared with each other if the networks have significantly different sizes 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 a holistic set of properties including the number of new author entries, exits, publications, and network connectivity measures, and perhaps even through graphical representations.

A discussion on the empirical extension or scope conditions is warranted. It is worth noting that collecting comprehensive data on authors, their collaborations, and funding sources, as was done in this paper for the black liquor gasification research network, can be significantly more challenging in larger and more diverse research domains. These difficulties arise from the sheer scale of research networks and the complexities of data collection. While these challenges exist, there is also merit in arguing that the fundamental principles and findings from this study can provide insights across a broader spectrum of research domains. Research networks, by their nature, exhibit similar characteristics in terms of collaboration, coauthorship, and the dynamics of knowledge dissemination (Kyvick and Reymert, 2017).

More importantly, the small-world property and the strength of weak ties, as discussed in Granovetter's (1973) seminal work, are not exclusive to any specific field but rather represent inherent characteristics of human social networks. Since the study's emphasis on the strategic use of funding leverages these core properties, the findings drawn from the study can be considered a transferable concept. While the specifics of funding allocation may vary from one field to another, the underlying principle of prioritizing funding strategies that promote network expansion and collaboration can be applicable in diverse research contexts. It is essential, however, for future research to carefully consider the unique conditions and dynamics within each field to determine the extent to which these findings can be generalized. This would require empirical studies in different research areas to validate the applicability of the proposed funding strategies beyond the specific sample studied in this paper.

6. Limitations

One concern that remains is the potential endogeneity between funding and publications over time. For example, more funding could improve the output of publication, yet more publications may facilitate greater funding success for researchers. The data set, however, is largely over a single time period as we have almost no measures on prior publication history on this topic to explain the first round of funding we observe. Also, there are almost no sequences in funding (for example, we have information on which specific funds led to which publication, but no information on which set of publications led to additional funding). What we do have is funding as exogenous data and related publications thereafter. Yet as funding and publications are perhaps joint successes, this really does not matter substantially for our own results if network relationships predict funding and subsequent publications.

An additional limitation is that there may exist correlation among the three econometrics models. This really is the same problem as funding problem above. For future studies, researchers may attempt to simulate a flexible covariance matrix in an R3 space, so that covariance can change at different points in the distribution. Work on three dimensions, however, is already computationally challenging.

7. 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 are 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 a set of 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. Using this publication information, we construct a network. Each researcher is a network node, and each vertex (network edge) 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 breaks 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.

Three funding strategies were compared for this experiment and all funding policies provide 22% more funding to the network overall. Overall, the findings suggest the type of network matters, but the features of this network are particularly compelling, at least for research policy. Direct Optimization, on average, increases publications by 92% and researcher recruitment by 28%; the Smart Small World Rule increases publication rates by 114% and researcher recruitment by 31%; the Fairness Rule increases publication rates by 111% and researcher recruitment by 37%. Finally, providing no additional funding reduces publications and the number of researchers. This experiment suggests that research funding that strengthens overall research community connectivity generates the highest levels of research productivity and connectedness.

We bridge several bodies of literature in research funding and its impact on innovation and network topology. Within these fields, we make two specific contributions. First, we demonstrate the applicability of funding strategies to stimulate a research network to achieve desirable individual- and overall network-level outcomes. We employ node turnover as drivers of topological change as a novel approach. Second, we show that targeted funding can be used as a policy vehicle to inject, and in some cases, strengthen and enhance desirable properties, such as the small-world property, into a network.

The results of our simulation exercise suggest implications for decision making in relation to funding scientific work. Our findings underscore the importance of considering the demonstrated ability of researchers to expand their network by recruiting new researchers as a key metric in evaluating grant proposals. This is anchored on the simulation results, which show that funding strategies that prioritize researchers with a higher degree of connections and ability to recruit new researchers to their network produce better outcomes in terms of network productivity and connectivity than strategies that simply reward research productivity.

Given these findings, researchers writing grant proposal can highlight their track record of successfully recruiting new researchers in their prior grant projects. By explicitly highlighting their capacity to expand research networks, proposal writers not only align with the priorities of funding bodies to broaden research ecosystem but also enhance the competitiveness of their proposals. At the same time, funding bodies can suitably adjust their evaluation criteria for grant proposals to take into consideration the number of new authors a researcher has successfully introduced into their research network as a parameter. In doing so, funding decisions become not only about assessing the projected innovation and productivity output of a proposal, but also about promoting the development of robust, resilient and diverse research networks. In turn, these networks, as indicated by our study, can contribute to an overall increase in *both* research productivity and network connectivity.

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. Amati, V., Lomi, A., Mascia, D., & Pallotti, F. (2021). The co-evolution of organizational and network structure: The role of multilevel mixing and closure mechanisms. Organizational Research Methods, 24(2), 285-318.
3. Anderlini, L., & Ianni, A. (1996). Path dependence and learning from neighbors. Games and economic behavior, 13(2), 141-177.
4. Ayoubi, C., Pezzoni, M., & Visentin, F. (2019). The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? Research Policy, 48(1), 84-97.
5. Ayoubi, C., Pezzoni, M., & Visentin, F. (2021). Does it pay to do novel science? The selectivity patterns in science funding. Science and Public Policy, 48(5), 635-648.
6. Azoulay, P., Liu, C. C., & Stuart, T. E. (2017). Social influence given (partially) deliberate matching: Career imprints in the creation of academic entrepreneurs. American Journal of Sociology, 122(4), 1223-1271.
7. Bajpai, P. (2016). Pulp and paper industry: energy conservation. Elsevier.
8. Bajpai, P. (2008). Chemical Recovery in Pulp and Papermaking. Surrey: Pira International.
9. 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.
10. Baruffaldi, S., Visentin, F., & Conti, A. (2016). The productivity of science & engineering PhD students hired from supervisors' networks. Research Policy, 45(4), 785-796.
11. Baum, J. A., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from? Industrial and Corporate change, 12(4), 697-725.
12. Beaudry, C., & Allaoui, S. (2011, September). Impact of research funding on nanobiotechnology scientific production: Does concentration in a few universities make sense? In 2011 Atlanta Conference on Science and Innovation Policy (pp. 1-28). IEEE.
13. Beaudry, C., & Schiffauerova, A. (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. European Management Journal, 29(5), 362-376.
14. Bolton, P., & Dewatripont, M. (1994). The firm as a communication network. The Quarterly Journal of Economics, 109(4), 809-839.
15. Boorman, S. A. (1975). A combinatiorial optimization model for transmission of job information through contact networks. The bell journal of economics, 216-249.
16. Chen, H., Mehra, A., Tasselli, S., & Borgatti, S. P. (2022). Network dynamics and organizations: A review and research agenda. Journal of Management, 48(6), 1602-1660.
17. 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.
18. Cowan, R., & Jonard, N. (2001). Knowledge creation, knowledge diffusion and network structure. In Economics with heterogeneous interacting agents (pp. 327-343). Springer, Berlin, Heidelberg.
19. Cowan, R., & Jonard, N. (2003). The dynamics of collective invention. Journal of Economic Behavior & Organization, 52(4), 513-532.

20. Cross, R. L., & Parker, A. (2004). The hidden power of social networks: Understanding how work really gets done in organizations. Harvard Business Press.
21. Dahlander, L., & McFarland, D. A. (2013). Ties that last: Tie formation and persistence in research collaborations over time. Administrative science quarterly, 58(1), 69-110.
22. 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.
23. den Hamer, P., & Frenken, K. (2021). A network-based model of exploration and exploitation. Journal of Business Research, 129, 589-599.
24. Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An experimental study of search in global social networks. science, 301(5634), 827-829.
25. Doreian, P., & Mrvar, A. (1996). A partitioning approach to structural balance. Social networks, 18(2), 149-168.
26. Ellison, G. (1993). Learning, local interaction, and coordination. Econometrica: Journal of the Econometric Society, 1047-1071.
27. Ellison, G., & Fudenberg, D. (1995). Word-of-mouth communication and social learning. The Quarterly Journal of Economics, 110(1), 93-125.
28. 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.
29. Feeney, M. K., & Bozeman, B. (2008). Mentoring and network ties. Human Relations, 61(12), 1651-1676.
30. 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.
31. Folkstad, J., & Hayne, S. C. (2011, January). Visualization and analysis of social networks of research funding. In 2011 44th Hawaii International Conference on System Sciences (pp. 1-10). IEEE.
32. Goyal, S., & Janssen, M. C. (1997). Non-exclusive conventions and social coordination. journal of economic theory, 77(1), 34-57.
33. Granovetter, M. S. (1973). The strength of weak ties. American journal of sociology, 78(6), 1360-1380.
34. Hargadon, A. (2003). How breakthroughs happen: The surprising truth about how companies innovate. Harvard Business Press.
35. He, J., & Fallah, M. H. (2009). Is inventor network structure a predictor of cluster evolution? Technological forecasting and social change, 76(1), 91-106.
36. Hilbert, M., Oh, P., & Monge, P. (2016). Evolution of what? A network approach for the detection of evolutionary forces. Social Networks, 47, 38-46.
37. Kim, L., Smith, S., Dahlander, L., & McFarland, D. A. (2022). Networking a career: Individual adaptation in the network ecology of faculty. Social Networks.
38. Kingsley, G., & Malecki, E. J. (2004). Networking for competitiveness. Small business economics, 23, 71-84.
39. Kingsley, G., Bozeman, B., & Coker, K. (1996). Technology transfer and absorption: an 'R & D value-mapping' approach to evaluation. Research policy, 25(6), 967-995.

40. Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. Advances in neural information processing systems, 14.
41. Kogut, B., & Walker, G. (2001). The small world of Germany and the durability of national networks. American sociological review, 317-335.
42. Leenders, R. T. A. (1996). Evolution of friendship and best friendship choices. Journal of Mathematical Sociology, 21(1-2), 133-148.
43. Lerner, J., & Lomi, A. (2020). The free encyclopedia that anyone can dispute: An analysis of the micro-structural dynamics of positive and negative relations in the production of contentious Wikipedia articles. Social Networks, 60, 11-25.
44. 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.
45. Mannak, R. S., Markus, A., Meeus, M. T., Raab, J., & Smit, A. C. (2023). Network dynamics and its impact on innovation outcomes: R&D consortia in the Dutch water sector. Social Networks, 74, 62-70.
46. Mannak, R. S., Markus, A., Meeus, M. T., Raab, J., & Smit, A. C. (2023). Network dynamics and its impact on innovation outcomes: R&D consortia in the Dutch water sector. Social Networks, 74, 62-70.
47. McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
48. Milgram, S. (1967). Six degrees of separation. Psychology Today, 2, 60-64.
49. Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. The American economic review, 81(5), 1408-1418.
50. Nakao, K., & Romney, A. K. (1993). Longitudinal approach to subgroup formation: Re-analysis of Newcomb's fraternity data. Social Networks, 15(2), 109-131.
51. Porter, A. L., Garner, J., & Crowl, T. (2012). Research coordination networks: evidence of the relationship between funded interdisciplinary networking and scholarly impact. Bioscience, 62(3), 282-288.
52. Porter, C. M., Woo, S. E., Allen, D. G., & Keith, M. G. (2019). How do instrumental and expressive network positions relate to turnover? A meta-analytic investigation. Journal of Applied Psychology, 104(4), 511.
53. Powell, W. V. (1990). Neither Market Nor. Research in organizational behavior, 12, 295-336.
54. Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. Administrative science quarterly, 116-145.
55. Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. American journal of sociology, 110(4), 1132-1205.
56. Prell, C., & Feng, K. (2016). The evolution of global trade and impacts on countries’ carbon trade imbalances. Social Networks, 46, 87-100.
57. Prell, C., & Lo, Y. J. (2016). Network formation and knowledge gains. The Journal of Mathematical Sociology, 40(1), 21-52.

58. Rawlings, C. M., McFarland, D. A., Dahlander, L., & Wang, D. (2015). Streams of thought: Knowledge flows and intellectual cohesion in a multidisciplinary era. Social Forces, 93(4), 1687-1722.
59. Robins, G., Pattison, P., & Woolcock, J. (2005). Small and other worlds: Global network structures from local processes. American Journal of Sociology, 110(4), 894-936.
60. Rumsey-Wairepo, A. (2006). The association between co-authorship network structures and successful academic publishing among higher education scholars. Brigham Young University.
61. Scatà, M., Di Stefano, A., La Corte, A., Liò, P., Catania, E., Guardo, E., & Pagano, S. (2016). Combining evolutionary game theory and network theory to analyze human cooperation patterns. Chaos, Solitons & Fractals, 91, 17-24.
62. Schiffauerova, A., & Beaudry, C. (2011). Star scientists and their positions in the Canadian biotechnology network. Economics of Innovation and New Technology, 20(4), 343-366.
63. Schiffauerova, A., & Beaudry, C. (2012). Collaboration spaces in Canadian biotechnology: A search for gatekeepers. Journal of Engineering and Technology Management, 29(2), 281-306.
64. 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.
65. Snijders, T. A. (1998). Methodological issues in studying effects of networks in organizations. Computational & Mathematical Organization Theory, 4(2), 205-215.
66. Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics. Annual review of statistics and its application, 4, 343-363.
67. Snijders, T. A., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. Social networks, 32(1), 44-60.
68. Snijders, T., & Doreian, P. (2010). Introduction to the special issue on network dynamics. Social Networks, 32(1).
69. US Energy Information Administration. (2019). Increases in electricity generation from biomass stop after a decade of growth. Available at: https://www.eia.gov/todayinenergy/detail.php?id=39052
70. 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.
71. Van Rijnsoever, F. J., Van Weele, M. A., & Eveleens, C. P. (2017). Network brokers or hit makers? Analyzing the influence of incubation on start-up investments. International Entrepreneurship and Management Journal, 13, 605-629.
72. Veloz, T., Razeto-Barry, P., Dittrich, P., & Fajardo, A. (2014). Reaction networks and evolutionary game theory. Journal of mathematical biology, 68, 181-206.
73. Verspagen, B., & Duysters, G. (2003). In press. The small worlds of strategic technology alliances. Technovation.
74. Wang, M. (2009). Stimulative Policies for a Dynamic Network of Bioenergy Researchers: Identifying Strategic Collaborations [Doctoral dissertation, Texas Tech University].
75. Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013). Exponential random graph models for multilevel networks. Social networks, 35(1), 96-115.
76. Watts, D. J. (2004). Small worlds: the dynamics of networks between order and randomness (Vol. 36). Princeton university press.

77. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. nature, 393(6684), 440-442.
78. Weesie, J., & Flap, H. D. (1990). Social networks through time. Isor.
79. 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.
80. Zhang, L., & Guler, I. (2020). How to join the club: Patterns of embeddedness and the addition of new members to interorganizational collaborations. Administrative Science Quarterly, 65(1), 112-150.
81. Zhao, S. X., Tan, A. M., Yu, S., & Xu, X. (2018). Analyzing the research funding in physics: The perspective of production and collaboration at institution level. Physica A: Statistical Mechanics and its Applications, 508, 662-674.

Tables and Figures

Table 1: Summary Statistics

	Poisson Model	Multinomial Logit	Logit Model
Dependent variable	0.426 (0.566)	2.170 (0.838)	0.478 (0.499)
Funding	0.085 (0.063)	0.446 (0.413)	0.446 (0.413)
First degree	11.501 (8.448)	26.740 (17.571)	26.740 (17.571)
Second degree	59.832 (20.389)	76.450 (26.718)	76.450 (26.718)
Shared coauthors	21.435 (106.051)	53.695 (29.714)	53.695 (29.714)
New authors	10.167 (50.757)	50.689 (29.082)	50.689 (29.082)
Funding square	1.121 (1.586)	3.692 (9.246)	3.692 (9.246)
First degree x second degree	775.411 (785.536)	2,424 (2,090)	2,424 (2,090)
N	*8,001*	*682*	*682*

Table 2: Models to estimate productivity and connectivity outcomes

	Poisson Model	Multinomial Logit Model		Logit Model
	Outcome: Productivity	Outcome: Entry		Outcome: Exit
		$y = 1$	$y = 2$	
Intercept	-1.941*** (0.211)	-1.065*** (0.017)	-1.703*** (0.014)	2.387*** (0.581)
Funding	0.236*** (0.025)	-0.021 (0.161)	0.103*** (0.054)	-0.082*** (0.009)
First degree	0.075*** (0.016)	-0.074** (0.031)	0.102*** (0.030)	-0.048*** (0.018)
Second degree	-0.002 (0.003)	0.007 (0.007)	0.007 (0.005)	-0.017*** (0.008)
Shared coauthors	0.447*** (0.078)	0.200 (0.128)	-0.486*** (0.054)	-0.966*** (0.188)
New authors	0.092** (0.040)	-0.242* (0.127)	0.492*** (0.055)	0.803 (0.686)
Funding square	-0.006*** (0.001)	-0.015 (0.014)	-0.003 (0.002)	0.001 (0.006)

First degree x second degree	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	0.001 (0.001)
N	*8,001*	*682*		*682*
Pseudo R-squared	*0.3879*	*0.0851*		*0.3559*
Log-likelihood	*-1271.0526*	*-1167.8766*		*-289.1522*
LR Chi-squared	*3404.36****	*225.54****		*179.98****

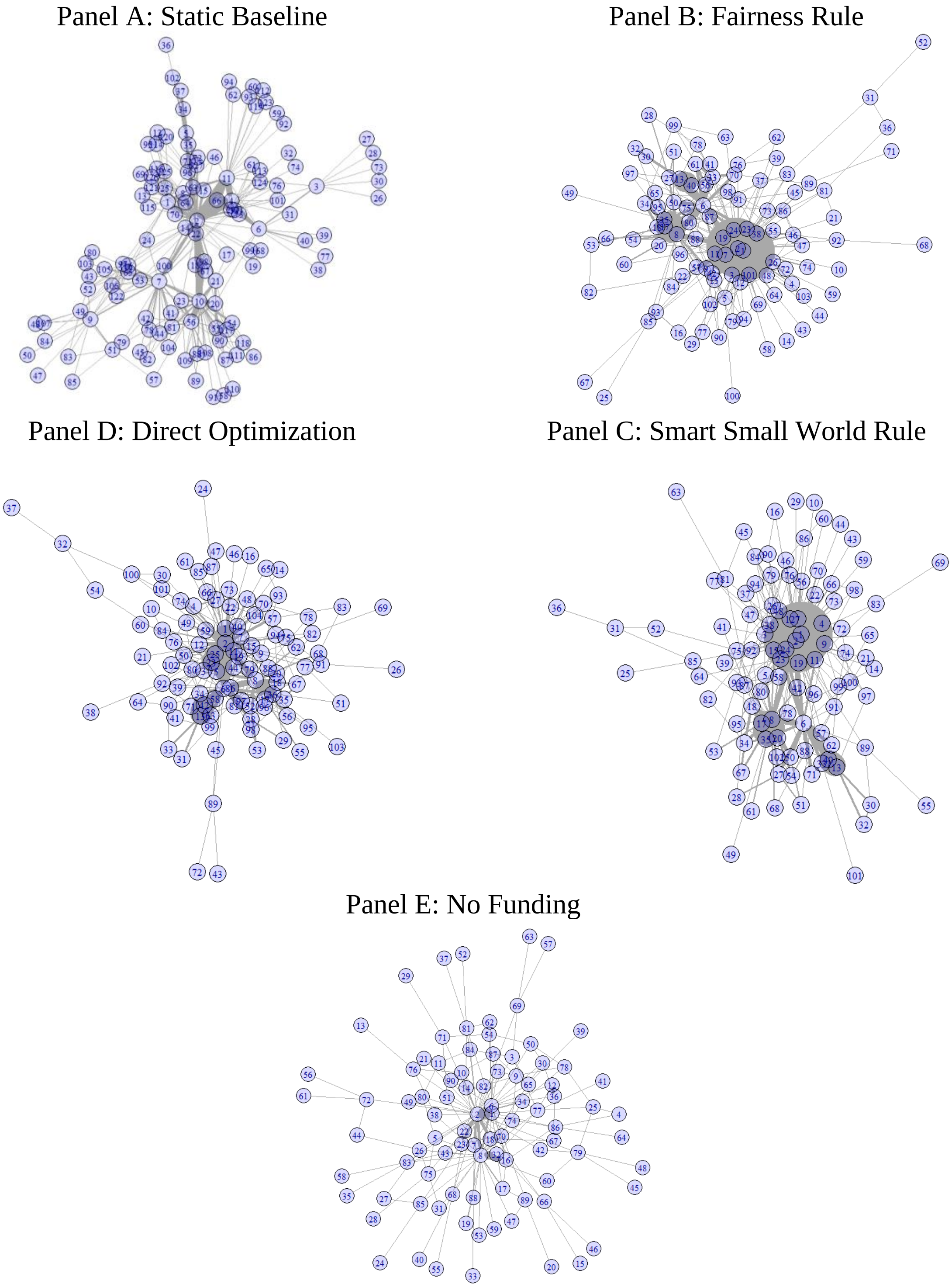
*Note: Standard errors are in parentheses. *p < 0.1, **p < 0.05, ***p < 0.01.*

Table 3: Summary Diagnostics

	Avg. # of publications	Avg. # of authors	Avg. shortest pathlength	Clustering coefficient	Small world index
Static Baseline	126 (4.81)	114 (5.49)	3.898 (0.308)	0.442 (0.022)	0.113 (0.011)
Fairness Rule	266 (8.17)	156 (4.24)	2.897 (0.796)	0.499 (0.020)	0.172 (0.025)
Smart Small World Rule	270 (7.11)	149 (3.43)	3.073 (0.589)	0.500 (0.029)	0.163 (0.016)
Direct Optimization	242 (5.93)	146 (3.01)	2.952 (0.545)	0.480 (0.014)	0.163 (0.019)
No Funding	116 (20.28)	90 (18.26)	2.992 (0.674)	0.357 (0.121)	0.119 (0.052)

Note: The network diagnostics are of the final network. Standard deviations in parenthesis.

Figure 1: Network Evolution Under Alternative Funding Rules



Note: Figures show the Static Baseline network, and the empirical networks after the final funding round. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

Appendix A: Supplementary Material

1. Data for Regression Models

This section of the Appendix covers details about how the data for the three regression models is obtained from the raw data of publication information.

1.1 Poisson Model (Productivity Model)

The data for the Poisson model is obtained from the maximum combination of 127 researchers, equivalent to (127*126/2=) 8,001 units of observation, with each unit representing every unique combination of researchers. We create all variables for each units, including the dependent variable of the number of publications shared by every researcher pair (for e.g., if authors A and B coauthored 5 paper, then the dependent variable is 5), and independent variables including funding (x_1), first degree (x_2), second degree (x_3), shared coauthors (x_4), new authors (x_5), funding square (x_6), and interaction of first degree and second degree (x_7), as shown in the following Table 1.

Table 1: Data for Poisson model

Combination	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1, 2	5	0.125	3	6	3	0	0.0156	18
1, 3	3	0.125	9	32	3	0	0.0156	288
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1, 127	1	0	8	34	0	0	0	272
2, 3	1	0	13	71	0	0	0	923
2, 4	2	0	15	65	0	0	0	975
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2, 127	3	0	13	71	0	0	0	923
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
126, 127	1	0	35	68	0	0	0	2380

Note: 8,001 observations

The nonnegative dependent variable y here is a count variable, which can take on a relatively few possible discrete values, including 0, so we can use Poisson model to see the effect of the explanatory variables on y and E$[y]$.

$$E(y|x) = \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7)$$

$$P(y = h|x) = \frac{\exp[-\exp(x\beta)]\,[\exp(x\beta)]^h}{h!}, \quad h = 0, 1, 2, \dots$$

Independent variables are funding (x_1), first degree (x_2), second degree (x_3), shared coauthors (x_4), new authors (x_5), funding square (x_6), and interaction of first degree and second degree (x_7).

1.2 Multinomial Logit Model (Entry Model)

The data for the Entry Model is obtained by pairing all author combinations for each publication (see Table 2). It is different from the Poisson model in that here, only authors who share a publication get paired (since pairing is by every publication), while in the Poisson model, every single author is paired, even if they share no publication. For example, if paper 1 has four authors: A, B, C, D, there will be six (4*3/2) possible author pairs.

Here, our objective is to determine how many new nodes will enter the network. The dependent variable here is determined by two new variables called "new" and "ongoing". If a specific paper includes a new author, then the value of "new" is equal to 1, or 0 otherwise. If a new researcher has more than one publication, then the value of "ongoing" for this person is equal to 1, or 0 otherwise. Because some new nodes will also exit after they enter, it is possible to have three categories for the dependent variable:

Choice 1: no new authors enter, y=1 (new = 0 & ongoing = 0), so the dependent variable for all combinations in this paper is 1.

Choice 2: new authors enter but only publish once, y=2 (new = 1 & ongoing = 0). This indicates that this paper recruits a new researcher, and the number of publications by this researcher is equal to 1. The value of the dependent variable for all combinations in this paper will be equal to 2.

Choice 3: new authors enter and they continue to publish more than once, y=3 (new = 1 & ongoing = 1). The value of the dependent variable for all combinations in this paper will be equal to 3.

Based on the three categories of dependent variable, we use a multinomial logit model to simulate how many new permanent authors enter the network.

$$P(y = h|x) = \frac{\exp(x\beta)}{1 + \exp(x\beta)}, \quad h = 1, 2, 3, \dots$$

All independent variables are the same as those in the Poisson model.

Table 2: Data for Multinomial Logit model

Paper	Combination	New	Ongoing	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Paper 1: A, B, C, D, E	A, B	0	0	1	0.125	3	6	3	0	0.0156	18
	A, C	0	0	1	0.125	9	32	3	0	0.0156	288
	A, D	0	0	1	31	74	0	0	0	0	2294
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	D, E	0	0	1	0	13	71	0	0	0	923
Paper 2: A, B, C	A, B	1	0	2	0	15	65	0	0	0	975
	A, C	1	0	2	0	12	72	0	0	0	864
	B, C	1	0	2	0	13	71	0	0	0	923
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Paper 126: A, B, D, E	A, B	1	1	3	0	35	68	0	0	0	2380
	A, D	1	1	3	0	8	34	0	0	0	272
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	D, E	1	1	3	0	8	34	0	0	0	272

Note: 682 observations

1.3 Logit Model (Exit Model)

The data for the Exit Model is obtained the same way as the Entry Model (see Table 3). If the paper is the last paper for any of authors in the combination, then the dependent variable is equal to 1, or 0 otherwise. For example, in Table 3, if the first paper is author A's last paper, then all combination of authors that includes author A in this specific paper will have a dependent variable equal to 1.

The objective is to estimate how many old nodes will exit the original network using a logit model to see if a connection will break, and then estimate the probability that any pair will disconnect. All independent variables are the same as those in the previous two models.

$$P(y = h|x) = \frac{\exp(x\beta)}{1 + \exp(x\beta)}, \quad h = 0, 1, 2, \dots$$

We then calculate the mean probability of each researcher in a pair to exit the network. As long as we know a connection is broken, we compare the two authors' probability of exiting and determine which of the two will exit the network.

Table 3: Data for Logit model

Paper	Combination	New	Ongoing	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Paper 1: A, B, C, D, E	A, B	0	0	1	0.125	3	6	3	0	0.0156	18
	A, C	0	0	1	0.125	9	32	3	0	0.0156	288
	A, D	0	0	1	31	74	0	0	0	0	2294
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	D, E	0	0	1	0	13	71	0	0	0	923
Paper 2: A, B, C	A, B	1	0	2	0	15	65	0	0	0	975
	A, C	1	0	2	0	12	72	0	0	0	864
	B, C	1	0	2	0	13	71	0	0	0	923
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Paper 126: A, B, D, E	A, B	1	1	3	0	35	68	0	0	0	2380
	A, D	1	1	3	0	8	34	0	0	0	272
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	D, E	1	1	3	0	8	34	0	0	0	272

Note: 682 observations

2. Static Baseline Simulation

This section of the Appendix covers details about the initial simulation process to create the Static Baseline network.

2.1 Poisson Model (Productivity Model)

We first calculate $(P(x = h), h = 0, 1, 2, \ldots, 19, 20)$ where h is the number of actual publications. Table 4 presents the probability matrix for each combination. 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, and then sum all simulated numbers and obtain a total connection of 688 for the whole network.

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. This process is repeated 1,000 times to eventually obtain a distribution of total publications (Figure 1).

Figure 1: Distribution of Publications in Static Baseline

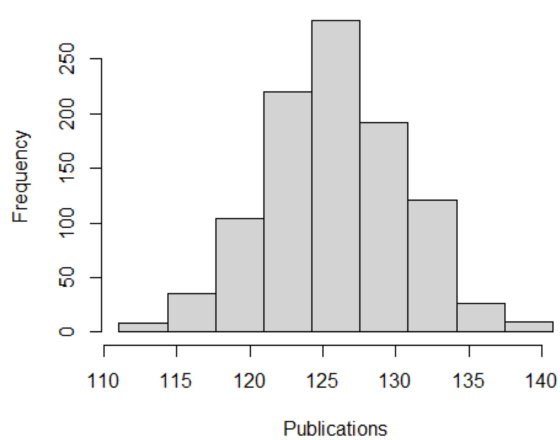


Table 4: Probability matrix for Poisson model

Combination	$\hat{y} = \lambda$	$P(y = 0)$	$P(y = 1)$	$P(y = 2)$	…	$P(y = 20)$
1, 2	0.07282	0.92976	0.06771	0.00246	…	1.089E-06
1, 3	0.07691	0.92597	0.07122	0.00274	…	1.349E-06
⋮	⋮	⋮	⋮	⋮		⋮
1, 127	0.02522	0.97509	0.02459	0.00031	…	1.644E-08
2, 3	0.03445	0.96613	0.03329	0.00057	…	5.671E-08
2, 4	0.02445	0.97583	0.02387	0.00029	…	1.455E-08
⋮	⋮	⋮	⋮	⋮		⋮
2, 127	0.02718	0.97318	0.02645	0.00036	…	2.214E-08
⋮	⋮	⋮	⋮	⋮		⋮
126, 127	0.02398	0.97630	0.02445	0.97580	…	1.346E-08

Note: 8,001 observations

2.2 Multinomial Logit Model (Entry Model)

Next, we simulate the number of new author recruitment into the network. We first calculate the probability of recruiting a new permanent researcher for each combination (y=3), 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 author connection does not recruit a new coauthor, and vice versa. This is presented in Table 5. The same duplication problem exists as in the previous case, and the 136 new connections from the first iteration can be similarly adjusted to estimate 25 new people entering the network.

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

Again, 1,000 iterations of the simulation yields the distribution of total author recruitment.

Table 5: Connection simulation for Entry model

Combination	$P(y = 3)$	Random number	Enter
A, B	0.07796	0.52314	0
A, C	0.25344	0.23421	1
A, D	0.63785	0.78945	0
⋮	⋮	⋮	⋮
			Total: 136

2.3 Logit Model (Exit Model)

Finally, we simulate the number of author exits from the network. To do this, we first estimate the probability that a connection is broken (y=1), and then draw a uniform random number to compare with the estimated probability. If the random number is lower than the probability, then this connection is predicted to break (table 6). We also calculate each author's probability of exit (Tables 7 and 8); so, when we identify a connection that breaks, we are able to compare the two authors' individual probability of exit and decide which author of the two will exit the system.

Using 1,000 iterations of the simulation, we obtain the distribution of author exits, the average of which is 38. The 1,000 simulations of author entries yielded an average total of 25 new author recruits. So, the Static Baseline network has 114 authors (13 fewer than the original network), the distribution of which is presented in Figure 2.

Figure 2: Distribution of Authors in Static Baseline

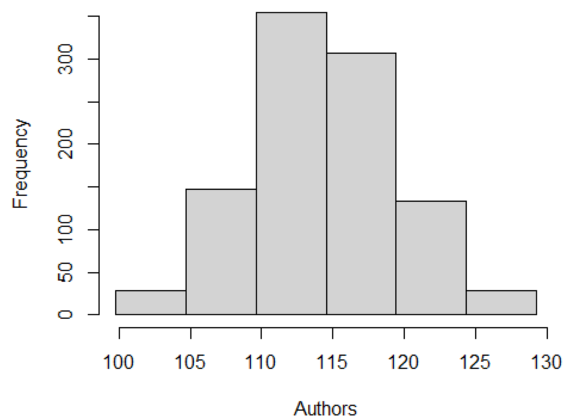


Table 6: Connection simulation for Exit model

Combination	$P(y=1)$	Random number	Connection break	Single probability	Exit
A, B	0.65426	0.52361	1	$P(A) > P(B)$	A
A, C	0.44658	0.23245	1	$P(A) > P(B)$	A
A, B	0.56124	0.78965	0	-	None
⋮	⋮	⋮	⋮	⋮	⋮
			Total: 204		

Table 7: Probability of exit by author pair

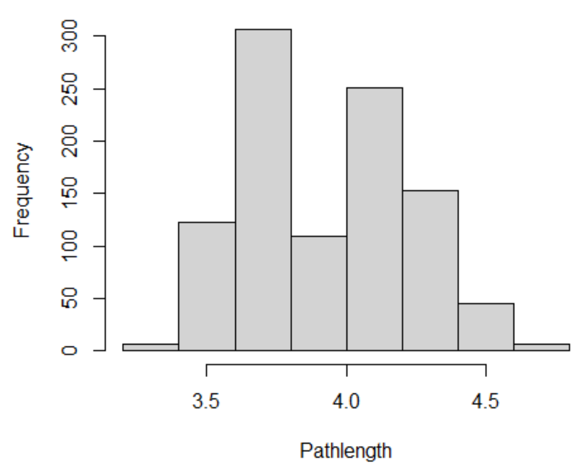
Combination	$P(y = 1)$
1, 2	0.1
2, 5	0.2
3, 7	0.3
1, 2	0.5
2, 5	0.4
⋮	⋮

Table 8: Author's probability of exit

Single prob.	$P(y = 1)$
1	$\frac{0.1 + 0.5}{2}$
2	$\frac{0.1 + 0.2 + 0.5}{3}$
3	…
5	…
7	…

Finally, the multinomial and logit regressions allow us to determine the number of new authors recruited by each author pair and pair breaks at each iteration. Based on this information, after every iteration, we attach the new authors into the network, and also adjust for the author who leave. This creates a new network after every iteration of the simulation, and we calculate the average shortest pathlength of the entire network at each iteration. A total of 1,000 simulations yields a distribution of average shortest path lengths, presented in Figure 3.

Figure 3: Distribution of Average Shortest Pathlength in Static Baseline

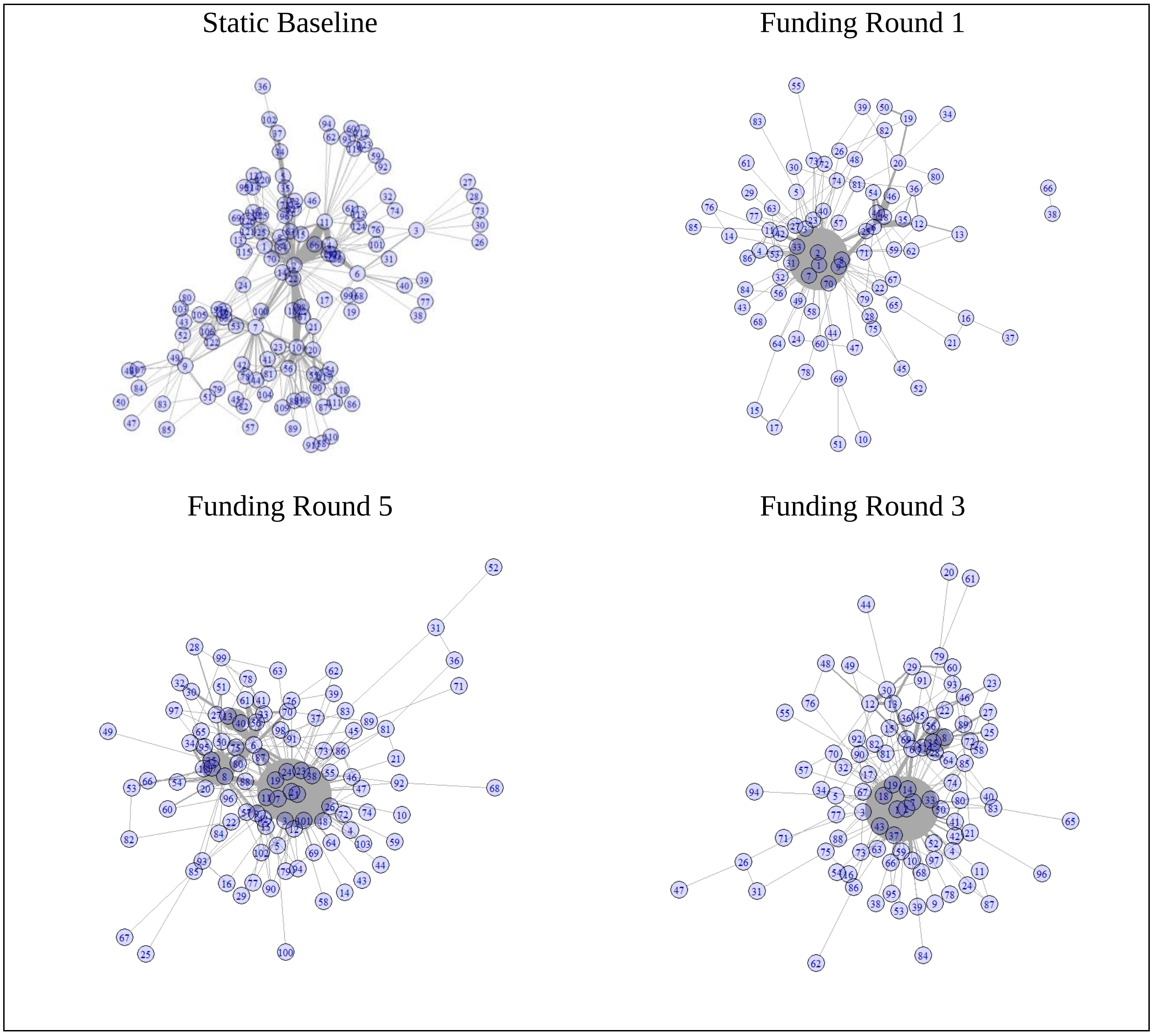


The shortest pathlength is formally defined as: $l_G = \frac{1}{n(n-1)} \sum_{i \neq j} d\left(v_i, v_j\right)$ where G is a graph with vertices V, $d(v_1, v_2)$ denote the shortest distance between v_1 and v_2, and n is the number of vertices in G.

Appendix B: Figures

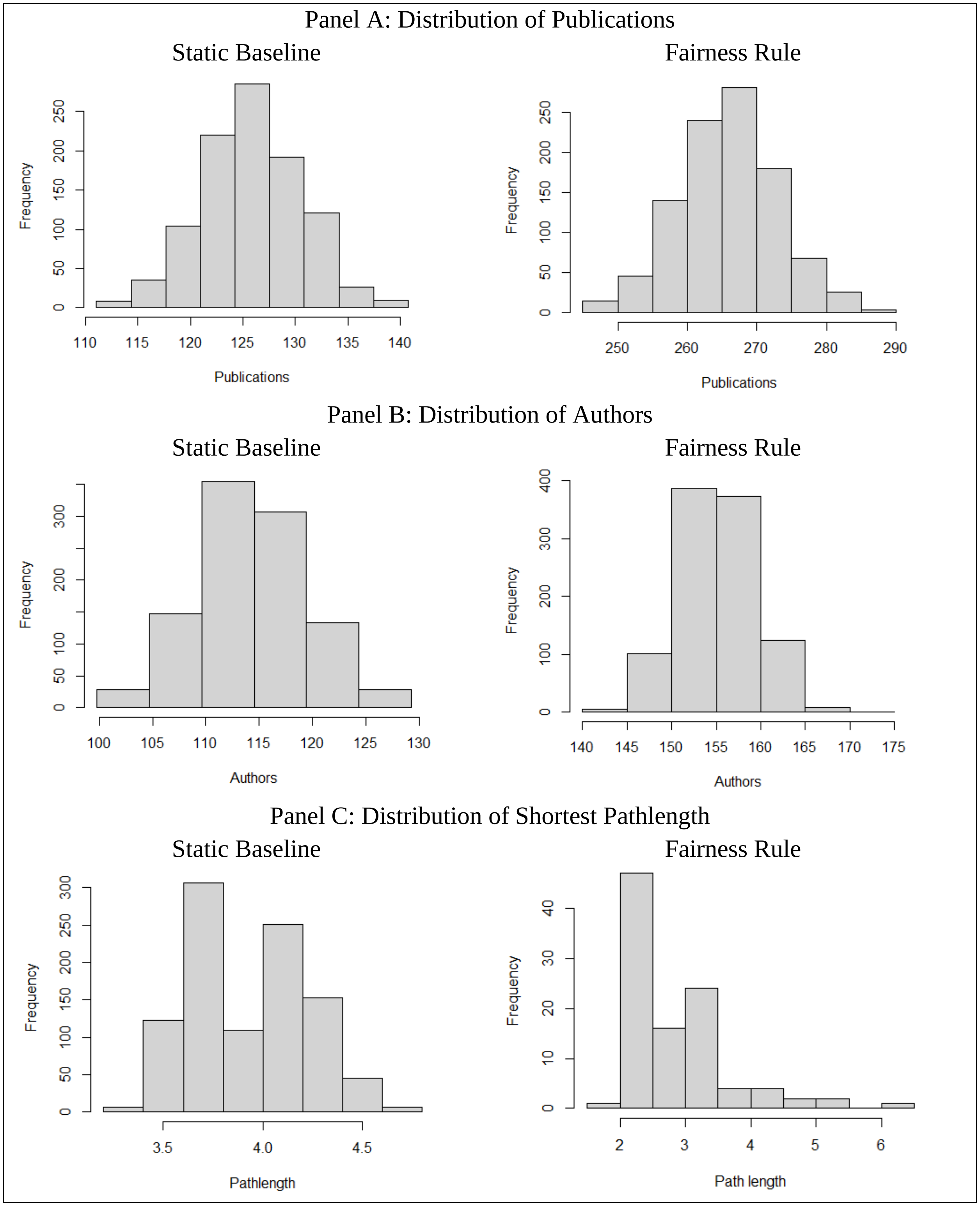
This section of the Appendix presents the (i) network diagrams as they evolve from the Static Baseline through funding rounds 1, 3, and 5; and (ii) distributions of the outcomes (number of publications, number of authors, and shortest pathlength) of the Static Baseline and the final network after the fifth funding round.

Figure 1: Network Evolution Under Fairness Rule



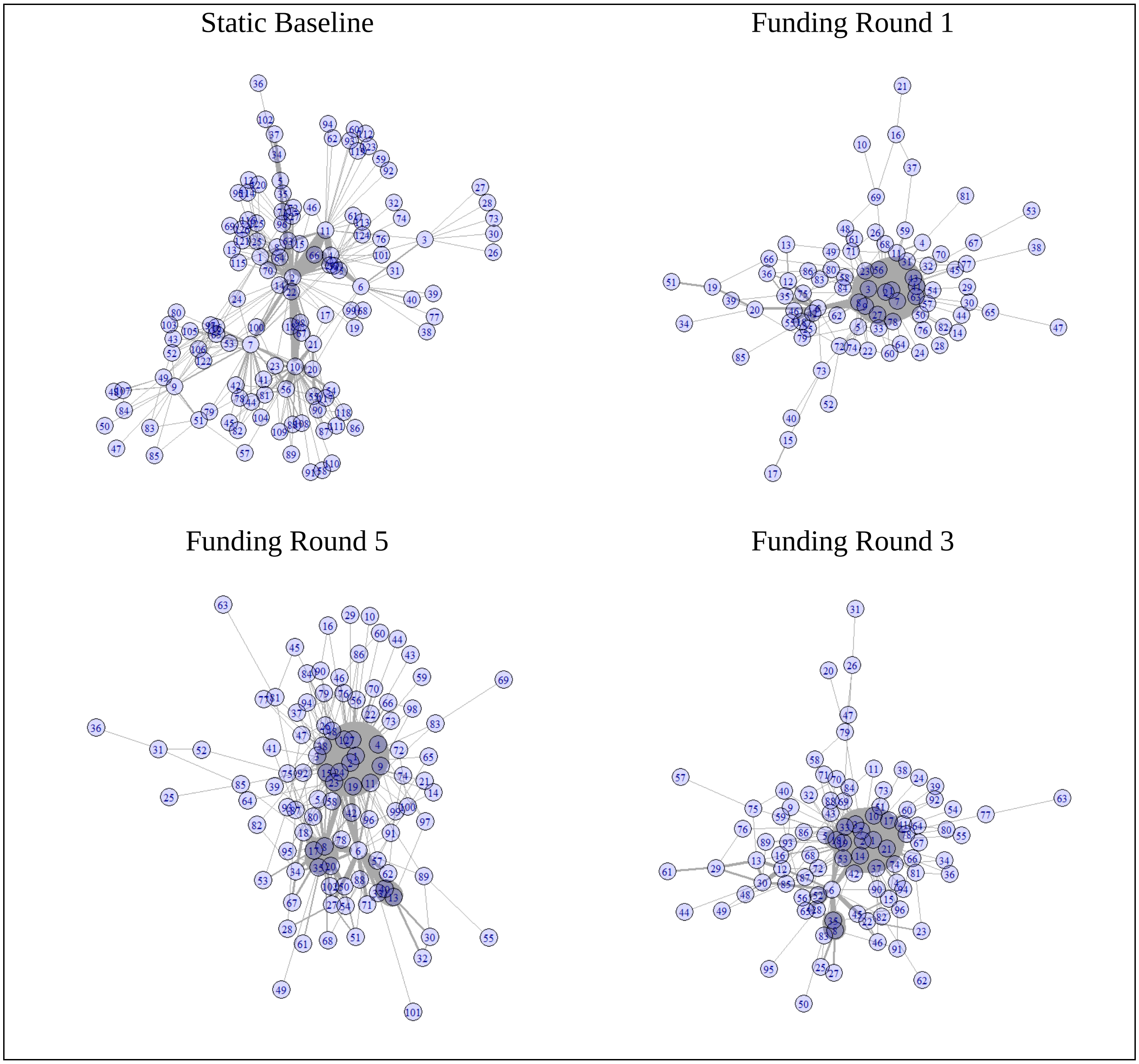
Note: Figures show the static baseline network, and the evolution of the network under the Fairness Rule. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

Figure 2: Distribution of Outcomes in Fairness Rule



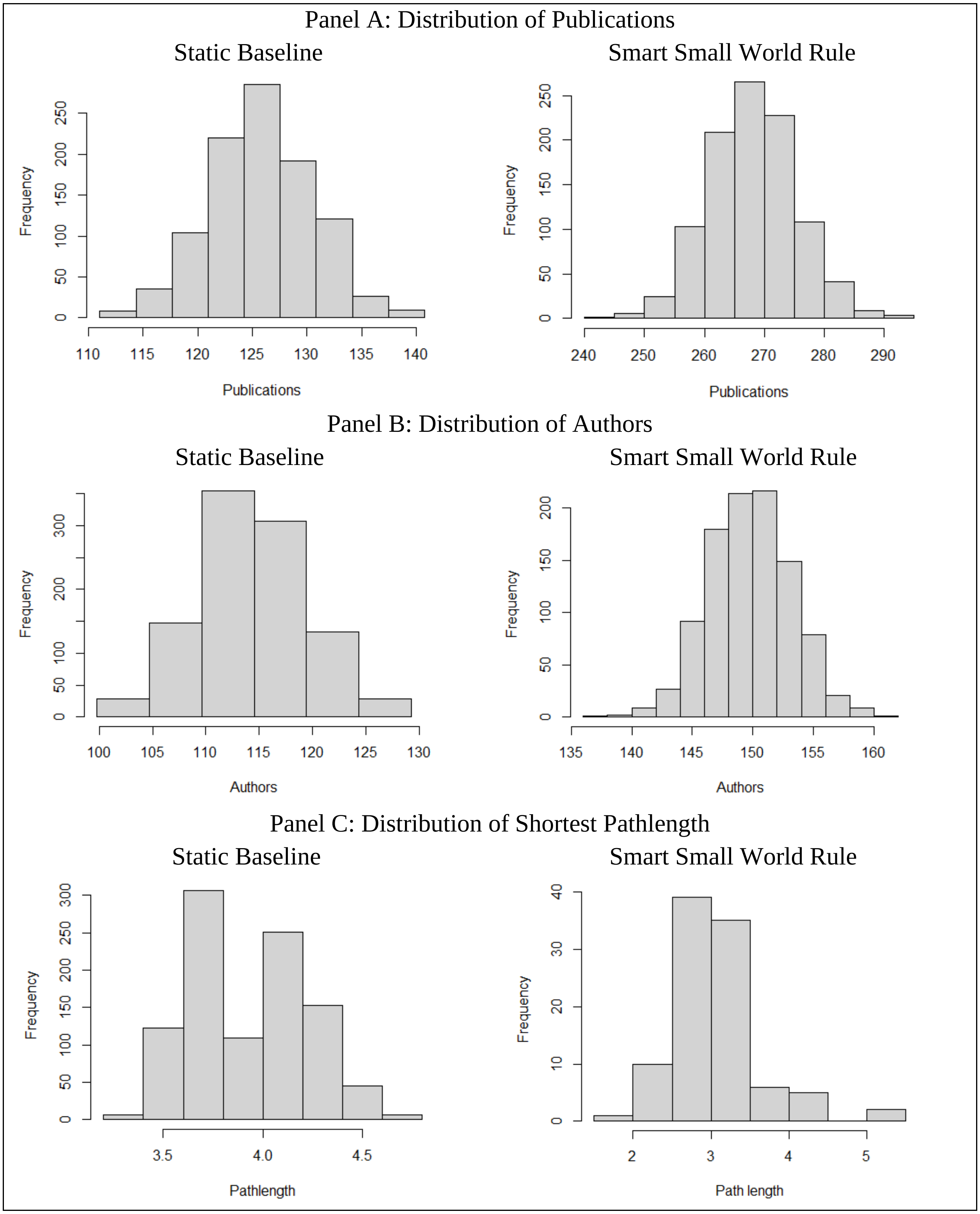
Note: Figures show distributions of number of publications, number of authors, and shortest pathlength of static baseline simulation and the final network simulation after fifth funding round under Fairness Rule.

Figure 3: Network Evolution Under Smart Small World Rule



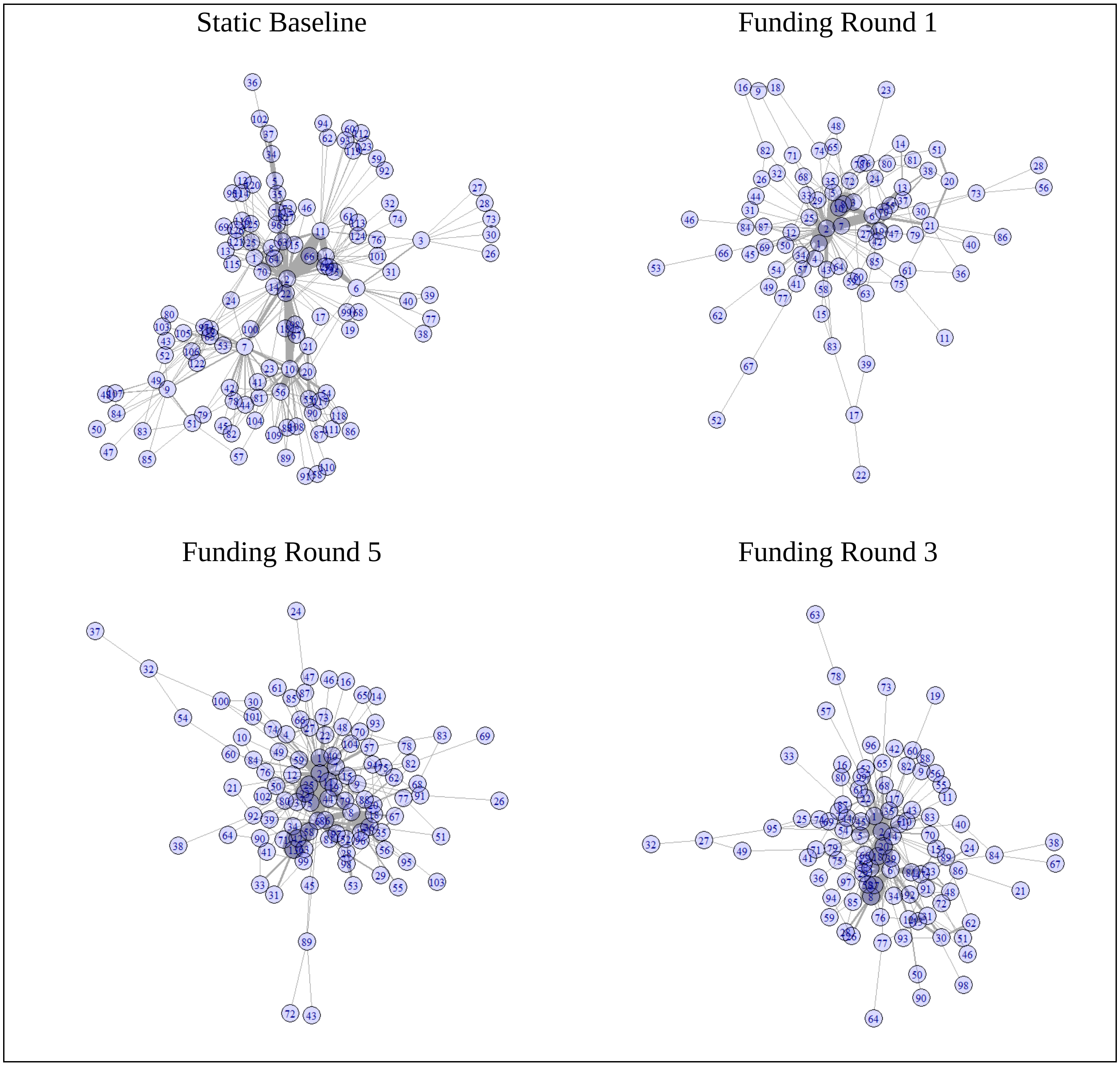
Note: Figures show the static baseline network, and the evolution of the network under the Smart Small World Rule. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

Figure 4: Distribution of Outcomes in Smart Small World Rule



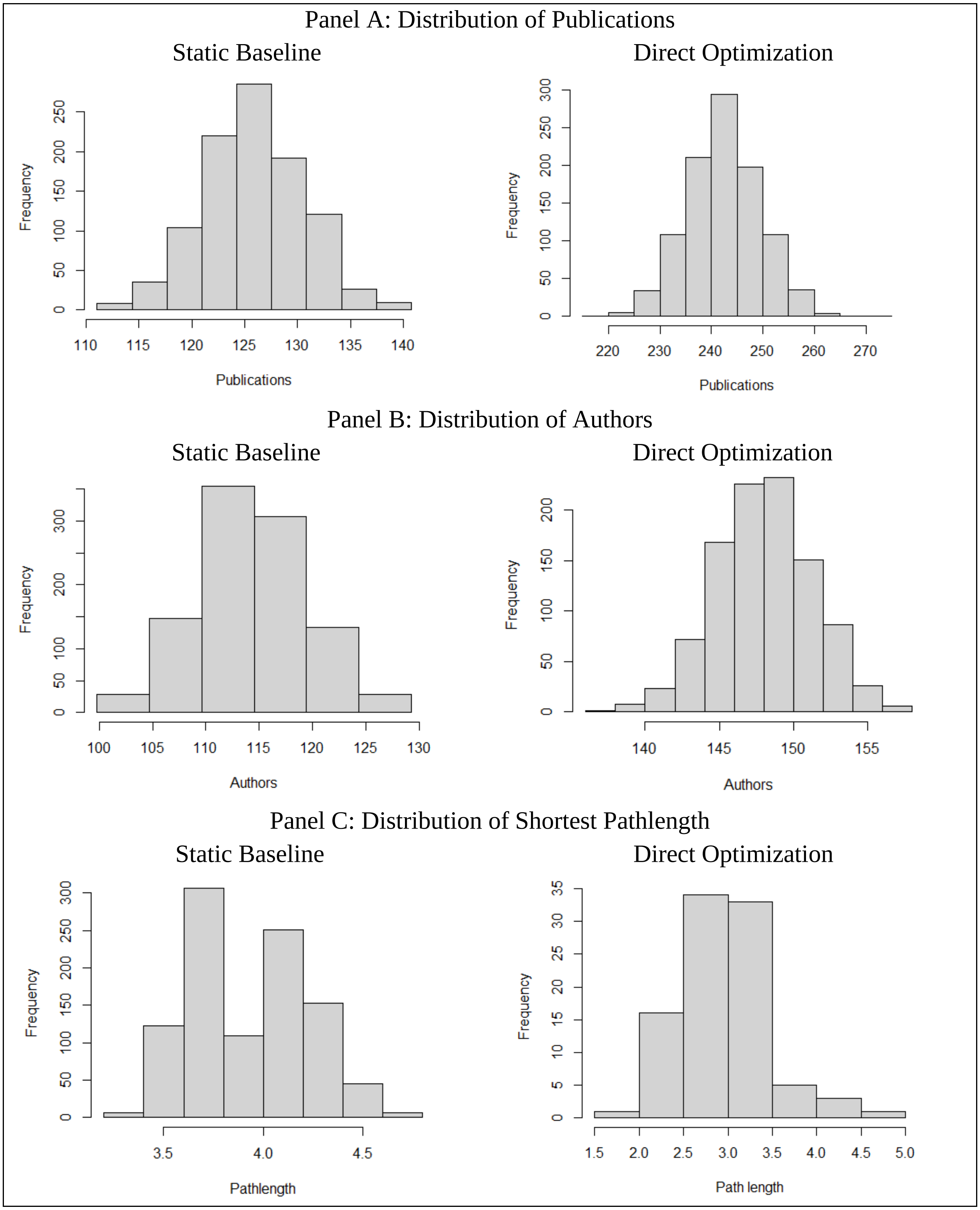
Note: Figures show distributions of number of publications, number of authors, and shortest pathlength of static baseline simulation and the final network simulation after fifth funding round under Smart Small World Rule.

Figure 5: Network Evolution Under Direct Optimization



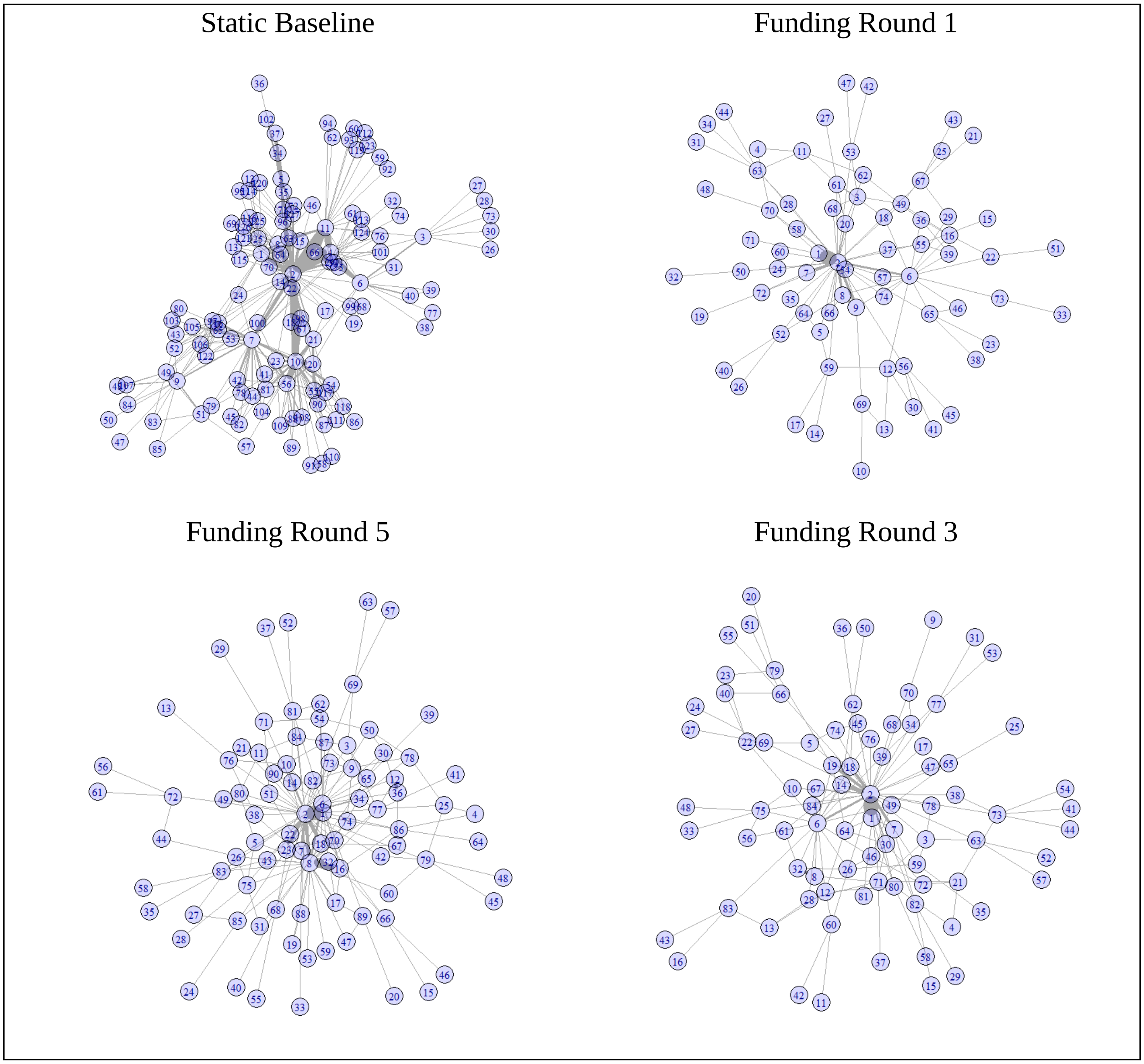
Note: Figures show the static baseline network, and the evolution of the network under Direct Optimization. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

Figure 6: Distribution of Outcomes in Direct Optimization



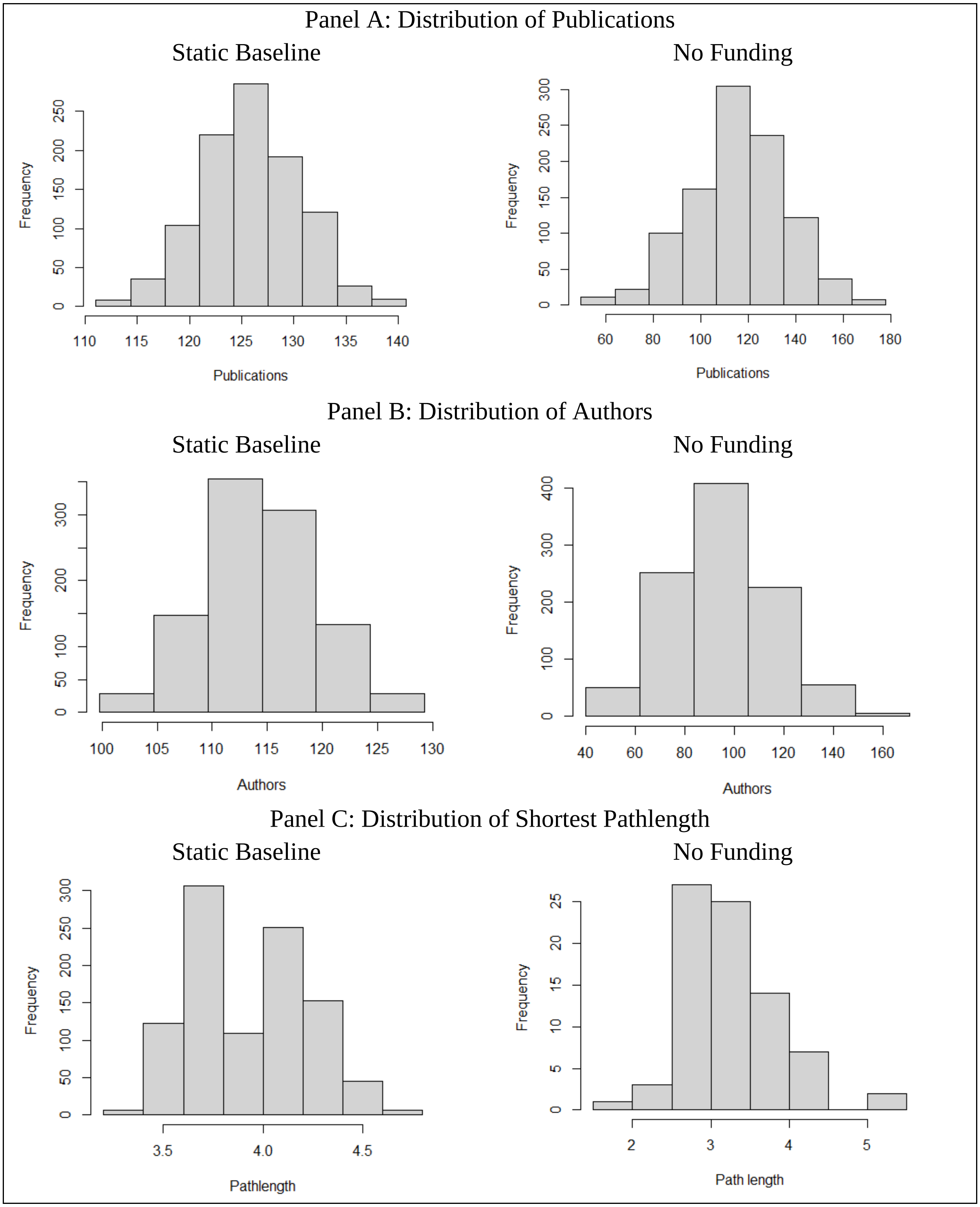
Note: Figures show distributions of number of publications, number of authors, and shortest pathlength of static baseline simulation and the final network simulation after fifth funding round under Direct Optimization.

Figure 7: Network Evolution Under No Funding



Note: Figures show the static baseline network, and the evolution of the network under No Funding. Each node represents an author, and the vertices connect every author pair who publish at least once. The thickness of the vertices represents the strength of the connection, defined by the number of publications shared between any two nodes.

Figure 8: Distribution of Outcomes in No Funding



Note: Figures show distributions of number of publications, number of authors, and shortest pathlength of static baseline simulation and the final network simulation after fifth funding round under No Funding.

Appendix C: Tables

This section of the Appendix presents the summary network diagnostics of the Static Baseline through funding rounds 1, 3, and 5.

Table 1: Network diagnostics by funding rounds under Fairness Rule

	Static Baseline	Funding round 1	Funding round 3	Funding round 5
Avg. # of publications	126	174	235	266
Avg. # of authors	114	122	140	155
Avg. shortest pathlength	3.898	3.465	3.120	2.897
Clustering coefficient	0.442	0.487	0.502	0.499
Small world index	0.113	0.141	0.161	0.172

Table 2: Network diagnostics by funding rounds under Smart Small World Rule

	Static Baseline	Funding round 1	Funding round 3	Funding round 5
Avg. # of publications	126	168	209	242
Avg. # of authors	114	115	139	148
Avg. shortest pathlength	3.898	3.568	3.195	2.952
Clustering coefficient	0.442	0.430	0.471	0.480
Small world index	0.113	0.121	0.147	0.163

Table 3: Network diagnostics by funding rounds under Direct Optimization

	Static Baseline	Funding round 1	Funding round 3	Funding round 5
Avg. # of publications	126	179	242	268
Avg. # of authors	114	122	144	150
Avg. shortest pathlength	3.898	3.561	3.123	3.073
Clustering coefficient	0.442	0.476	0.482	0.500
Small world index	0.113	0.134	0.154	0.163

Table 4: Network diagnostics by funding rounds under No Funding

	Static Baseline	Funding round 1	Funding round 3	Funding round 5
Avg. # of publications	126	106	124	116
Avg. # of authors	114	115	102	90
Avg. shortest pathlength	3.898	3.412	3.105	2.992
Clustering coefficient	0.442	0.426	0.386	0.357
Small world index	0.113	0.125	0.124	0.119