NEW ENGLAND
COMPLEX SYSTEMS INSTITUTE

Science for the challenges of society.

Corporate Competition

A Self-Organizing Network

Dan Braha
Blake Stacey
Yaneer Bar-Yam

REPORT 2010-3-01: March 15, 2010

24 Mount Auburn Street
Cambridge, MA 02138
T 617-547-4100
F 617-661-7711
www.necsi.edu

Corporate Competition: A Self-Organized Network

Dan Braha

New England Complex Systems Institute, Cambridge, Massachusetts and University of Massachusetts, Dartmouth, Massachusetts

Blake Stacey and Yaneer Bar-Yam

New England Complex Systems Institute, Cambridge, Massachusetts

(Dated: March 19, 2010)

A substantial number of studies have extended the work on universal properties in physical systems to complex networks in social, biological, and technological systems. In this paper, we present a complex networks perspective on interfirm organizational networks by mapping, analyzing and modeling the spatial structure of a large interfirm competition network across a variety of sectors and industries within the United States. We propose a model that is able to reproduce experimentally observed characteristics of competition networks as a natural outcome of a minimal set of general mechanisms governing the evolution of competition networks. The model suggests that macro dynamical processes determine to a large extent the ecology of industry structure. There is an asymmetry between companies that are considered competitors, and companies that consider others as their competitors. All companies only consider a small number of other companies as competitors, however there are a few companies that are considered as competitors by many others. Geographically, the density of corporate headquarters strongly correlates with local population density, and the probability two firms are competitors declines with geographic distance. We construct these properties by growing a corporate network with competitive links using random incorporations modulated by population density and geographic distance. Despite randomness, the historical order of incorporation matters to network structure. Our new analysis, methodology and empirical results are relevant to various phenomena of organizational behavior, and have implications to research fields such as economic geography, economic sociology, and regional economic development.

Keywords: Organizational Networks; Interfirm Competition; Economic Geography; Social Networks; Spatial Networks; Network Dynamics

I. INTRODUCTION

In recent years, major advances have been made in understanding the structure and dynamics of real-world social, biological, and technological complex networks [1–4]. Complex networks theory has also contributed to organizational and managerial environments, where new theoretical approaches and useful insights from application to real data have been obtained [5–7]. Through theory and experiment, these studies have characterized the structural properties of such networks, their mechanisms of formation, and the way these underlying structural properties provide direct information about the characteristics of network dynamics and function. Of particular interest are scale-free networks where the degree (*i.e.,* the number of nodes adjacent to a node) is distributed according to a power law or a long right tail distribution. Such networks have characteristic structural features like "hubs", highly connected nodes [8], features which cause them to exhibit super-robustness against failures [9, 10] on the one hand, and super-vulnerability to deliberate attacks and epidemic spreading [11] on the other. Modeling real world large interfirm competition networks, which capture the coupling between economic units, is important to understanding the complex dynamics, robustness, and fragility of economic activity.

Here, we use network methodology to analyze and model the spatial structure of a large competition network, representing competitive interactions among firms within the United States. We find that the framework of geographic complex networks, mainly applied to natural and engineered systems, can be extended to capture the underlying structure and macro dynamics of interfirm competition, a system of heterogeneous economic units involved in strategic interaction. We extend the understanding of organizational problems by following the "empirics first" multidisciplinary approach [1, 3, 5, 6, 47, 48]. To study complex systems, comprising many interacting units, we first look for robust empirical laws that describe the complex interaction followed by theoretical models that help understand and reproduce the main properties of the real world system. Our study combines several empirical measurements of competition networks and theoretical models, which are then validated and informed by the actual measurements. We focus initially on four fundamental properties: node degree distributions, the spatial distribution of firms, the relationship between connectivity probability and geographic distance, and edge length distribution. We then propose a simple model where new firms are added to the 2D surface of the Earth, and become connected to existing firms following a preferential attachment rule that is also dependent on geographic distance. We show that the model is able to remarkably reproduce the observed measurements.

We represent corporate entities by network nodes, and we connect nodes using directed edges following the competitive relationships of the firms involved. A network-based approach to competition views competition as a property of observable social ties among identifiable organizations, and as including conscious recognition by the organizations about the relation (see, *e.g.,* [46]. (This is distinct from the view taken in organizational ecology, in which organizations may compete even though they lack awareness of one another's existence and therefore cannot take one another's actions into account [52]. This type of diffuse competition is defined as the combined effect of many organizations upon a given organization. Following this interpretation, the potential for two firms to compete is further defined in terms of the degree of intersection or overlap of their "fundamental niches" – the set of resources and environmental conditions that can sustain the functioning of firms [52–55].)

While most network-theory research has concerned nodes and connections without any reference to an underlying space, in many real-world networks nodes and links are embedded in a physical space. In such networks, the interactions between the nodes depend on the geometrical distance between nodes; often, edges tend to link nodes that are close neighbors. Examples include natural, engineered, and social networks [12] such as the physical arrangement of the Internet [13–16], road and airline networks [16–18], broadcast signaling networks [19], power grids [20], mobile communication networks [21], and neuronal networks [22]. In real-world systems, the probability that two nodes are connected has been seen to decrease as a power law [14, 21, 23] or an exponential [15] of the distance between them. Other research has characterized the geographical deployment of nodes in two or three-dimensional Euclidean space. For example, Yook *et al.* [14] and Lakhina *et al.* [15] have shown that in technologically developed countries the Internet demand (measured by router density) is proportional to the population density. Several models of spatial networks have been proposed in the complex systems literature among which include placing nodes on simple regular lattices that are either joined randomly depending on their distance or if their distance is less than a certain cutoff [24, 25]; combining network growth and preferential attachment modulated by distance selection mechanisms [14, 26, 27]; and generating geographic networks based on local optimization processes [16].

Concerns of geographic and social proximity are not unknown in the social sciences. In sociology, gravity-based models predict that the likelihood of a relationship is inversely proportional to the physical distance between two individuals [43, 44]. In the context of international economics, the gravity model of trade predicts trade-flow volumes and capital flows between two units to be directly proportional to the economic sizes of the units (using GDP data) and inversely proportional to the distance between them [45]. In economic geography, the gravity model was used to explain migration flows between countries, regions, or cities [46], and showed that movement of people between cities is proportional to the product of their population size and inversely proportional to the square of the distance between them.

Spatial networks have also been of interest to economic geographers, who considered networks as a means for understanding urban growth, geographical clusters, international trade, and globalization [28]. These efforts, however, have been mostly metaphorical and insufficiently formalized [29]. In sociology and organization theory, models of networks (including spatial networks) have largely focused on the factors that affect the dynamics of the formation of linkages between members of a network [30–34]. These empirical studies provide support for preferential attachment type of mechanisms [8] as an important driver of tie selection [34–37]. For example, the alliance behavior of multinational corporations suggests that firms will be more likely to have further alliances in the future with increasing experience and connectivity [33, 36], and an expanding network of interfirm alliances in American biotech exhibits preferential attachment [37]. Geography as a significant determinant of tie selection and network expansion has also been demonstrated. Empirical research illustrates that ties between firms, representing alliances, corporate board interlocks, or investments, are more likely when two firms are co-located [34, 37–39]. Moreover, studies show that geographical proximity affects the entry of firms in a network forcing them to locate in spatial proximity to industry agglomeration [40, 41].

In Section 2, we represent real-world data on corporate competition and headquarter location as a directed network in space. In Section 3, we report an asymmetry between the in-degree (number of corporations a firm is affected by) and out-degree (number of corporations a firm affects) distributions. Next, in Section 4, we report that the geographic arrangement of corporate headquarters strongly correlates with population density and that the probability two firms are competitors declines with geographic distance. In Section 5, we develop a single model for spatial network growth that yields both the degree distributions and geographic statistics of the empirical network. We conclude in Section 6 with implications for the field of economic geography.

II. CORPORATE GEOGRAPHIC COMPETITION NETWORK DATA

The competition network was reconstructed from information records provided by Hoover's – a large business research company that offers comprehensive business information through the Internet on corporations and organizations in over 600 industries. Within the detailed company records, information can be found on location type (headquarters or other); street, city and state address; financial information; industry codes; and competitors list. The competitors list was selected based on various information sources including public documents (*e.g.,* SEC filings), company websites, industry-specific trade and journals, and directly from the company themselves. In order to avoid problems of disjoint maps, we limit our study to firms with headquarters locations in the contiguous United States.

Note that the location of a firm's headquarters is not necessarily the location of all that firm's activities. A large corporation can have many local or regional offices (for example, Google is headquartered in Mountain View, California, but has branches in other US cities including Atlanta, Boston, Chicago, New York and Washington, D.C. [42]); a major industrial manufacturer can have its main corporate offices in one city and factories scattered elsewhere. We focus here on headquarters location, for which it is simplest to define a unique value for each firm. We will see that the statistics show headquarters location to be a meaningful quantity with considerable predictive power.

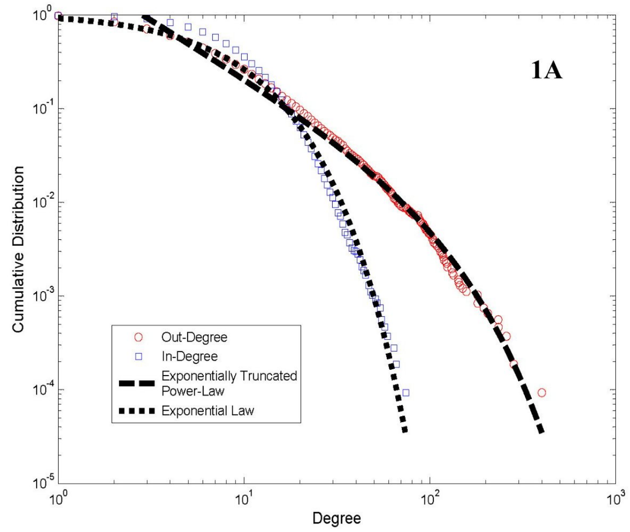
The competition network can be studied by several sampling methods [50]. Here, we use "snowball sampling" (*e.g.,* [51]) starting from a single node (company), we select all of the nodes directly linked from it, then the nodes linked from those selected in the last step, continuing until the desired number of nodes are sampled. Snowball sampling is a useful technique when relational data is not given explicitly. In our case, Hoover's maintains the corporate data in a website with the following structure: webpages correspond to different companies, and each webpage includes general data related to the company as well as a list of companies judged to be competitors. Thus, to construct the network, we perform snowball sampling. We begin with a company and collect its list of competitors; in a recursive fashion we traverse the list of competitors to other webpages, collecting more list of competitors and so on. This method could possibly generate a network that does not reflect (in a statistical fashion) the structure of the "real" network, because we start the sampling from a particular node. To eliminate this bias, one can construct different networks by starting the same snowball sampling method from different seed companies. After one computes several sampled networks G_1, G_2, G_3, $\ldots$, one takes the union of these networks (the set of nodes is the union of the set of nodes in G_1, G_2, G_3, $\ldots$, and duplicated arcs are excluded) to get a larger sampled network. This procedure can in principle generate a network with multiple connected components, which is indeed the case for our corporate competition network.

Our sampling of the business information site, combining sampled networks starting from companies whose main activities are in different industries, resulted in a directed network of 10753 companies and 94953 links. Taking the undirected version of the network, we find that it has three components: one giant components that includes 10718 nodes; and three small components with 2, 1, and 1 nodes. This occurs because the original sampled networks include companies around the globe, while the network analyzed is reduced to companies that operate within the United States. Consequently, despite the overlaps among sampled networks, reducing the network to include only United States companies yields disconnected nodes.

III. ANALYSIS OF IN- AND OUT- DEGREE DISTRIBUTIONS

A competition network can be considered as a directed graph with N nodes and L arcs, where there is an incoming arc to company v_i from company v_j if company v_i lists v_j as a competitor. There are 94,953 competition links in this directed network, with an average number of about 9 incoming (or outgoing) arcs connected to a node, and 40% of the competition links being reciprocal.

We compared (see Figure 1) the cumulative probability distributions $P_{\text{in}}(k)$ and $P_{\text{out}}(k)$ that a company has more than k incoming and outgoing links, respectively. The out-degree distribution can be described by a power law (the "scale-free" property) with a cutoff



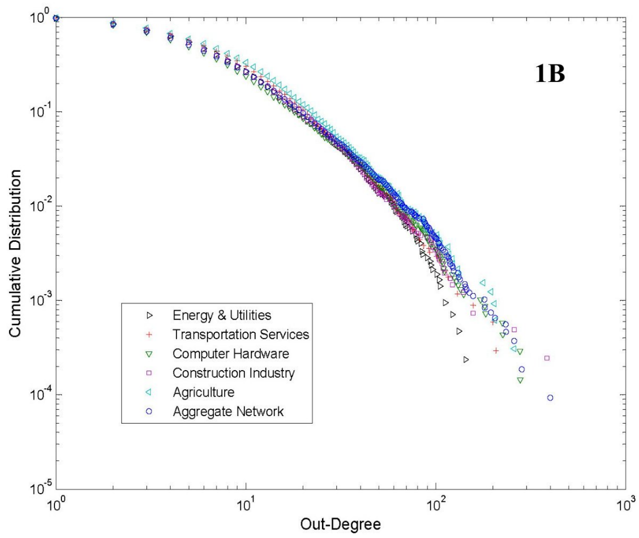


FIG. 1: The log-log plots of the cumulative distributions of incoming and outgoing links. **(A)** Aggregate competition network (10753 nodes, 94953 arcs). The out-degree distribution is characterized by an exponentially truncated power-law with $\gamma \approx 2.2$. The in-degree distribution is fitted by an exponential ($\exp -\frac{k}{7.1}$). **(B)** Several sampled networks. The networks were identified by snowball sampling, starting from a seed company whose main activity is in a particular industry (Energy & Utilities: 4248 nodes, 35789 arcs; Transportation Services: 3408 nodes, 31330 arcs; Computer Hardware: 6914 nodes, 56663 arcs; Construction Industry: 4094 nodes, 35376 arcs; Agriculture: 3245 nodes, 32758 arcs). We find that the out-degree distributions of all sampled networks are similar to each other as well as to the aggregate network, suggesting the robustness of the snowball sampling used here.

introduced at some characteristic scale $k^{-\gamma}e^{-k/k^*}$, with $k^* \approx 92$ and $\gamma \approx 2.2$. In contrast, the in-degree distribution is described by a fast decaying function, implying that companies with large incoming connectivity are practically absent. In this case, the in-degree distribution might be better fitted by an exponential as seen by the markedly curved-shaped behavior in Figure 1. Thus, while companies typically consider as competitors a small number of other companies, there are a few companies that are considered as competitors by many others.

Asymmetric in- and out- degree distributions have been found in other large complex networks [5, 48]. The connectivity of competition networks is important in constraining and determining many aspects of dynamical processes occurring on top of them, such as pricing decisions, strategic behavior, and firm performance. For example, it stands to reason that events and activities of central firms will tend to quickly propagate (due to the heterogeneous incoming connectivity) throughout the entire competition network, benefiting or impairing the vitality of the interconnected firms. This seems similar to ecological networks, where the loss of a keystone species could have large effects on the network [52].

We next examine the spatial characteristics of competition networks. The specific latitude and longitude of each company was obtained from its address using Yahoo's Geocoding Web Service, and the distance between two companies was calculated by using their geographical coordinates. In Figure 2, we compare the geographical deployment of companies with the population distribution in the contiguous U.S. Visual inspection of these maps suggests that the spatial distribution of companies is strongly correlated with the population distribution. We note that this kind of correlation was also observed for the geographical deployment of Internet routers [14, 15].

The patterns shown in Figure 2 can be explained as dynamic interplay between population growth and firm emergence. It is plausible that new ideas generated by entrepreneurs are in proportion to the size of the underlying population. New firms and job opportunities, in turn, will tend to increase the size of the local population (employed directly or indirectly by the firms), which will lead to further ideas, etc.

IV. ANALYSIS OF GEOGRAPHICAL DISTRIBUTIONS

The corporate competition spatial network enables us to relate competition and geographic distance. Figure 3 shows the probability $P(v_i \rightarrow v_j | d(v_i, v_j) = l)$ that two companies

2A. Geographical Deployment of Firms

2B. The 2000 Population Distribution in the Contiguous United States

FIG. 2: **(A)** Geographical deployment of companies with headquarters locations in the contiguous United States. The latitude and longitude of each company was obtained from its address using Yahoo's Geocoding Web Service. **(B)** Map of the 2000 Population Distribution in the United States (also referred to as the "Nighttime Map") produced by the U. S. Census Bureau as part of the 2000 Decennial Census. In this map, white dots coalesce to form the urban population concentrations; each white "dot" represents 7,500 people.

separated by a distance l are related by a competition link. As the distance l increases, we observe that the probability tends to decrease according to a power-law, indicating that geographic proximity increases the probability of competition. However, the fluctuations around the power-law behavior for distances larger than $\approx 1000\,\text{km}$ suggest that a model for the presence of competition needs to take into account both geography-dependent mechanisms and non-geographic processes.

The physical distance between nodes in geographic networks plays an important role

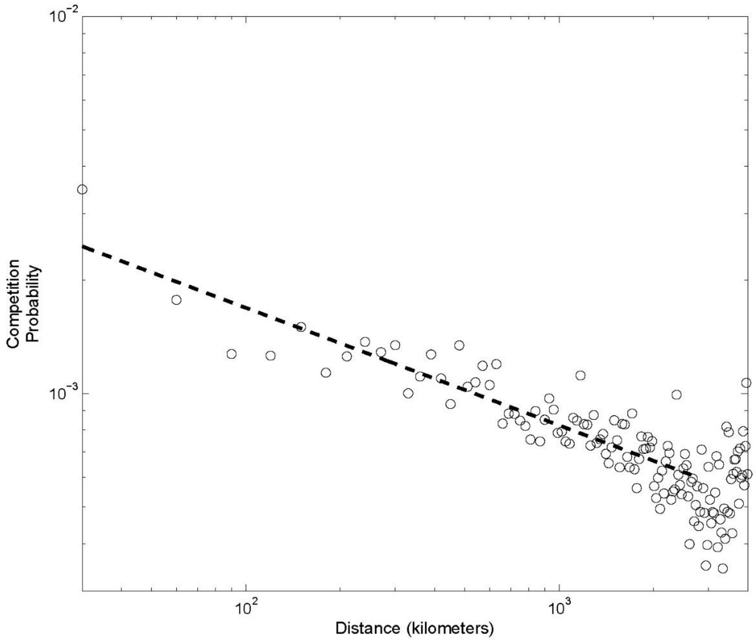


FIG. 3: The relationship between geographic distance and competition. The log-log plot shows the probability that two companies separated by a distance l are related by a competition link. The competition probability is fitted by a power-law $l^{-\beta}$ with $\beta \approx 0.3$. The probability was estimated from the proportion of pairs of connected companies separated by a distance l among all the total number of pairs of (connected or not) companies separated by a distance l (practically, link lengths with a resolution of 30 km are examined).

in determining the costs and benefits of communication and transport. As such, common to many geographic networks is a bias towards shorter links. The competition network analyzed here is of no exception (but perhaps for different reasons). We show in Figure 4 the cumulative probability distribution that the length of a link is greater than l kilometers. We find that the link length distribution can be well fitted by an exponential function and sharp subsequent decay.

The geographic nature of the competition network also has an effect on its *topological robustness.* In network theory, "robustness" refers to a network's ability to withstand attacks, such as random deletion of nodes or the targeted removal of highly-connected hubs. The effect of attacks is typically gauged by the change effected in the network's topological properties, such as the size of its largest component: a network which falls apart into many

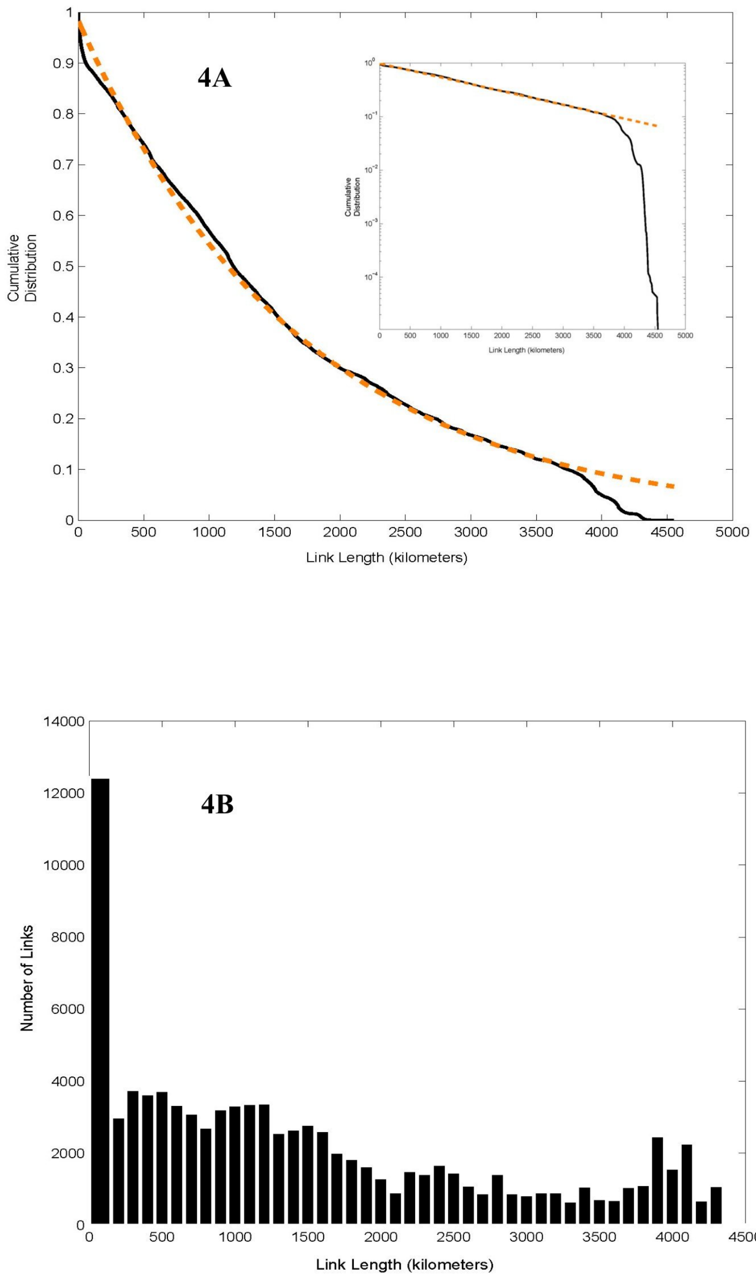


FIG. 4: **(A)** The cumulative probability distribution that the length of a link is greater than kilometers. The cumulative distribution is well fitted by an exponential $(\exp -\frac{l}{1688})$ with a sharp subsequent decay. The inset shows the corresponding semilog plot. **(B)** The histogram of the lengths of links. We see that the competition network has many very short links of length 100 km or less, a large portion of links of length 3800 km or less, and then an apparent smaller peak of longer links around 4000 km. Many of these longer links represent continent-wide distances.

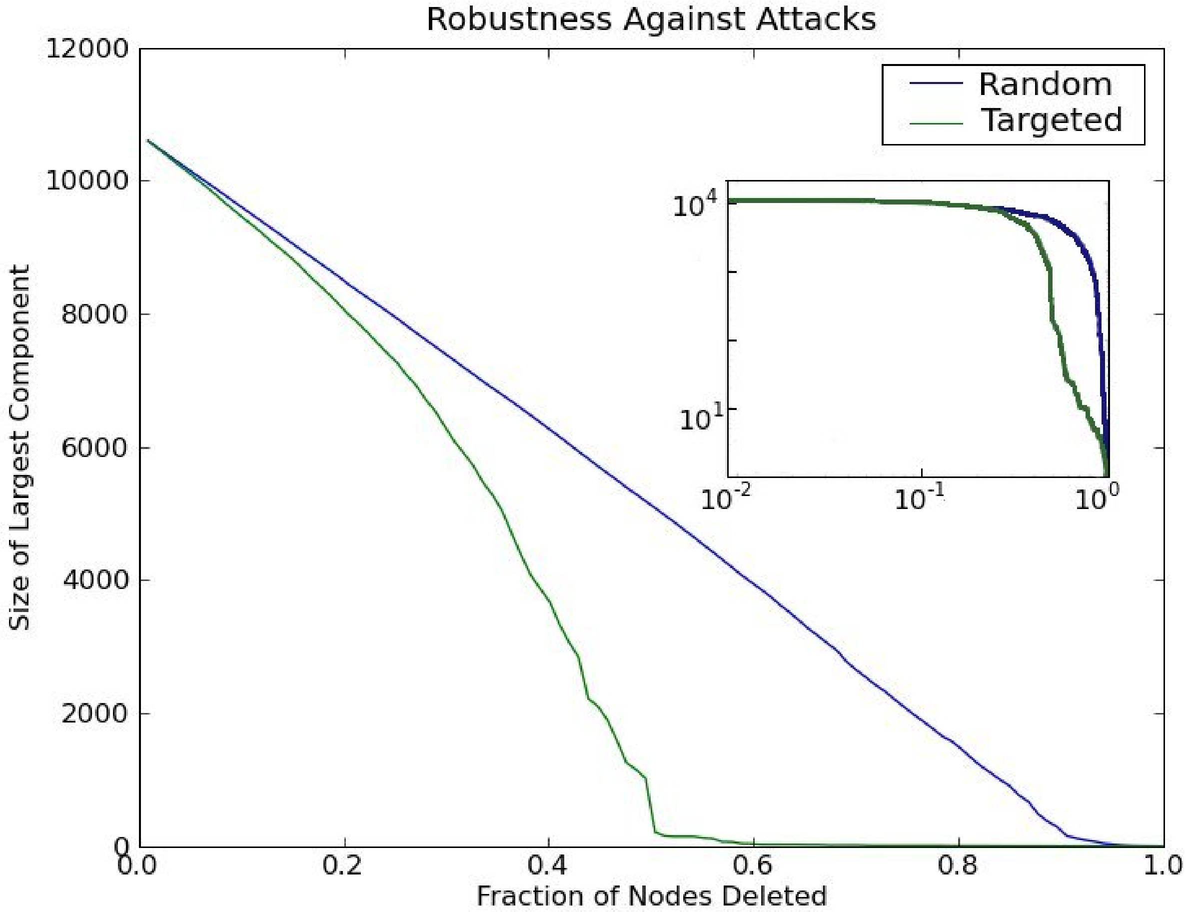


FIG. 5: Robustness of the competition network, as demonstrated by the deletion of nodes. Blue (upper) line shows the size of the largest component surviving as nodes are randomly deleted; green (lower) line shows the size of the largest component as nodes are deleted in order of decreasing degree. Higher-degree nodes are more "central" in that attacking them breaks apart the network more efficiently. Inset: the same data, plotted logarithmically.

small pieces upon the excision of a single node is fragile. Power-law networks grown by preferential attachment have been found to be resilient against random attacks, but weak against the targeted deletion of high-degree nodes. Here, we see that the degree distribution is not the only relevant factor in determining robustness; as Figure 5 shows, even after we delete thousands of nodes, the network does not dissolve into disconnected pieces.

V. MODELING CORPORATE SPATIAL COMPETITION NETWORKS

Although the competition network studied here represents data collected at particular time point, it is the result of a specific development path of network dynamics that involve firm entries and exit as well as the formation and dissolution of competition links. The heavy tail characteristic displayed by the out-degree distribution of the competition network (Figure 1) suggests that the evolution of competition networks could be governed by a preferential attachment rule [8]. On the other hand, Figures 3 and 4 show that firms compete according to the distance between them, suggesting that the formation of links is geometrical in nature. Moreover, the strong correlation between firm and population distributions (Figure 2) suggests that any model of competition should take into account the concentration of firms in highly populated areas. We therefore seek a model that considers the interplay between preferential attachment, geographic distance, and population density effects. A real understanding and modeling of competition networks should be able to reproduce experimentally observed characteristics of competition networks — such as the degree and link length distributions reported in Figures 1 and 4 — as natural outcome of a minimal set of general mechanisms governing the evolution of competition networks.

Network growth models including geographical distance of nodes [14, 25, 26] are a natural modeling approach for competition networks. We identify points on the curved surface of the Earth by their latitude and longitude coordinates, and compute geographic distances using the great circle distance between pairs of points on the surface of a sphere. We then superimpose on the map a grid consisting of two sets of parallel longitude and latitude lines, dividing the Earth's surface into squares (for our numerical simulations, we use high resolution data that consists of boxes of $0.0083° \times 0.0083°$). At each box, the population density is calculated from population data by dividing the population of each box by its area in square kilometers. In the following, firms are distributed on the Earth's surface by sampling from the population density distribution. We start with m_0 firms, each pair connected by a competition link, and at each subsequent step the network grows with the addition of new firms. For each new firm, m new directed competition links are created connecting it to firms already present in the system. The exponentially truncated power-law distribution of the outgoing connections (Figure 1) suggests the use of a nonlinear preferential attachment rule [1, 3], which generalizes the linear preferential attachment mechanism that results in a power-law degree distribution [8]. This, combined with the fact that the competition probability tends to decrease with geographic distance according to a power-law, offers the possibility that the growth of competition networks is governed by a nonlinear preferential attachment rule modulated by a link length dependent factor. More specifically, the firms j receiving the new links from firm i are chosen with probability proportional to $k_j^{\alpha} l_{ij}^{\beta}$, where k_j is the total degree of firm j; l_{ij} is the length in kilometers of the directed link from i to j; α and β are continuously varying parameters.

We have tested the validity of the above model by conducting several extensive computational experiments, and comparing the simulation results with the actual observations made from the competition network. In the simulations below, we have used a population density grid from the 2000 U.S. Census produced by the Columbia University Center for International Earth Information Network (CIESIN). The grid has a resolution of 30 arc-seconds (0.0083 decimal degrees), or approximately 1 square km. In all cases, we start with $m_0 = 8$ connected firms, and at each step of network growth a new firm with $m = 8$ directed competition links will be connected to firms already present in the system until the total number of firms reaches the actual number of firms in the sampled competition network $N = 10753$.

The model above offers a good flexibility for calibration since two parameters can be modified. Altering the value of α and β will influence the estimated spatial interactions. We have calibrated our model to correctly reproduce the experimentally observed degree and link length distributions characteristics reported in Figures 1 and 4, and have derived the best-fitted values of $\alpha = 0.85$ and $\beta = -0.3$ (henceforth called the Competition model). In addition, we examine three extreme cases of the competition network model: (1) Linear Preferential Attachment: $\alpha = 1$, $\beta = 0$; (2) Gravity I: $\alpha = 0$, $\beta = -1$; and (3) Gravity II: $\alpha = 0$, $\beta = -2$. The first case corresponds to the scale-free network model developed by Barabási and Albert [8] where an already present firm receives a new competition link according to a linear preferential attachment rule, that is, with probability proportional to its degree. The second and third cases reflect a variety of gravity models in social science that are based on the empirical principle that proximity in geographic (and social) space affects the likelihood of interaction [43–46].

Figures 6A and 6B compare the link length and out-degree distributions, respectively, generated by the above four models with the experimentally observed distributions shown in

Figures 1A and 4A. We note that both the link length and out-degree distributions of the real competition network deviates significantly from that produced by the Linear Preferential Attachment, Gravity I, and Gravity II models. The simulation results, however, of the Competition model are able to nicely reproduce the actual observations of the competition network, indicating that the Competition model gives a better characterization of the data than the three models specified above. The Competition model takes into account three effects: population density, preferential attachment, and geographic distance. The value of $\alpha = 0.85$ reflects a sublinear ($\alpha < 1$) tendency of preferential linking to firms with many competition links, which can result in a truncated power-law degree distribution [2, 5, 46], as indeed observed experimentally (Figure 1A). The value of $\beta = -0.3$ shows that the "friction of distance," or how rapidly interaction decreases as distance increases, is relatively small compared to the Gravity-based models.

In order to test for the effect of geographical distance bias on the competition network growth, we have held the "preferential attachment" parameter α at its optimal value 0.85 and have varied the "friction of distance" parameter β set at values 0, -0.3, -1, and -2. Figure 7A shows how a decrease in the value of β leads to a significant bias towards shorter links. However, Figure 7B shows that the out-degree distributions are almost not affected by the value of β, when α is set at the optimal value 0.85. This suggests that the parameter α has a strong effect on the out-degree distribution, and weak effect on the link length distribution.

Testing for the effect of preferential attachment on the competition network growth further corroborates this finding. This is done by experimenting with varying values of α, when β is set at the optimal value -0.3. As shown in Figures 8A and 8B, while the link length distributions for varying α are not changed and are similar to the actual distribution, the out-degree distributions deviate significantly for values of α that are different from the optimal value 0.85. These results imply that the actual link length distribution is determined to a large extent by the "friction of distance" parameter β, and weakly so by the "preferential attachment" parameter α. Overall, Figures 7–8 show once more that the Competition model provides a better characterization of the competition network than other combinations of α and β.

Finally, we test for the effect of population density on the competition network growth. To this end, we have set the values of α and β to their optimal values, and have chosen

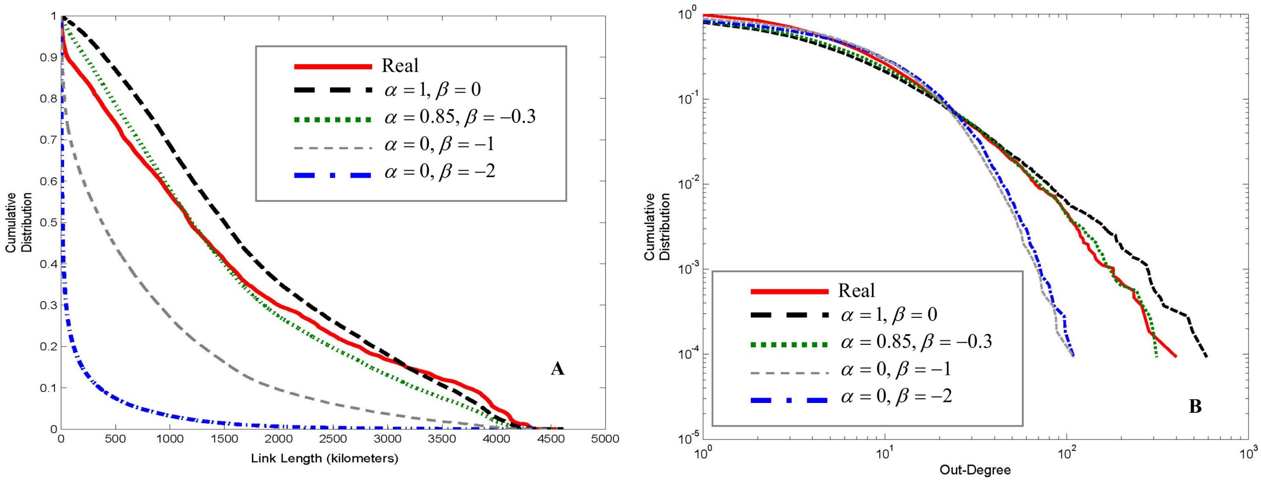


FIG. 6: Comparison between the actual competition network and simulation results of four network competition models, with respect to the cumulative link length **(A)** and out-degree **(B)** distributions. The models considered are: Linear Preferential Attachment (black dashed line, $\alpha = 1,\ \beta = 0$); Competition model (green dotted line, $\alpha = 0.85,\ \beta = -0.3$); Gravity I (gray dashed line, $\alpha = 0,\ \beta = -1$); Gravity II (blue dashed-dotted line, $\alpha = 0,\ \beta = -2$). Distributions corresponding to the actual competition network are shown as red solid lines. In the simulations, the contiguous United States range from to latitude, and from to longitude, and we have used a 3120 by 7080 population density grid from the 2000 U.S. Census with a box resolution of 30 arc-seconds (0.0083 decimal degrees), or approximately 1 square km. In all cases, firm location on the surface of the contiguous United States is determined by randomly sampling from the population density distribution. Once a box is sampled, the firm's latitude and longitude are uniformly and randomly located within the box. **(A)** Comparison of cumulative link length distributions. **(B)** Comparison of cumulative out-degree distributions. The out-degree distribution generated by the Preferential Attachment model (with linear preferential attachment and without geographical distance effect) is fitted by a power-law; the out-degree distribution corresponding to the Competition model is characterized by an exponentially truncated power-law; and both Gravity I and Gravity II (with only geographical distance effect) generate distributions that are well fitted by an exponential.

the location of firms based on two methods: (1) Location by Population Density (as in the Competition model): firms are distributed on the surface of the contiguous U.S. by sampling from the population density distribution; and (2) Random Location: firms are distributed randomly and uniformly on the surface of the contiguous U.S. Figure 9B shows that both firm

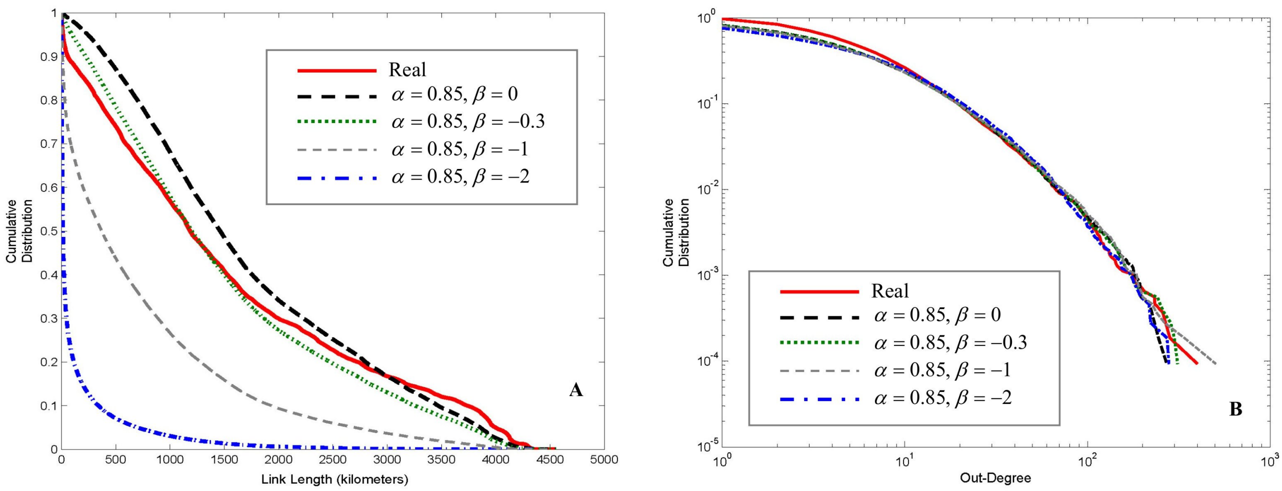


FIG. 7: The effect of the geographical distance bias on the link length **(A)** and out-degree **(B)** distributions. In the simulations, we have held the "preferential attachment" parameter α at its optimal value 0.85, firm location is determined by the population density distribution, and we have varied the "friction of distance" parameter β. Four network competition models are considered: Sublinear Preferential Attachment without Geographical Distance Bias (black dashed line, $\alpha = 0.85,\ \beta = 0$); Competition model (green dotted line, $\alpha = 0.85,\ \beta = -0.3$); Sublinear Preferential Attachment with Inversely Linear Distance Bias (gray dashed line, $\alpha = 0.85,\ \beta = -1$), and Sublinear Preferential Attachment with Inversely Square Distance Bias (blue dashed-dotted line, $\alpha = 0.85,\ \beta = -2$). Distributions of the actual competition network are shown as red solid lines.

placement schemes give similar results when comparing their out-degree distribution results with that of the observed data. However, as shown in Figure 9A, the link length distribution produced by the random location scheme deviates significantly from that produced by both the Competition model and actual competition network. In summary, Figures 6–9 provide good evidence that the structure of competition networks can be better explained by taking into account network dynamical growth, preferential attachment, geographical distance, and demographic factors such as population density. In particular, a simple model that is able to reproduce reasonably well the main observed features was proposed.

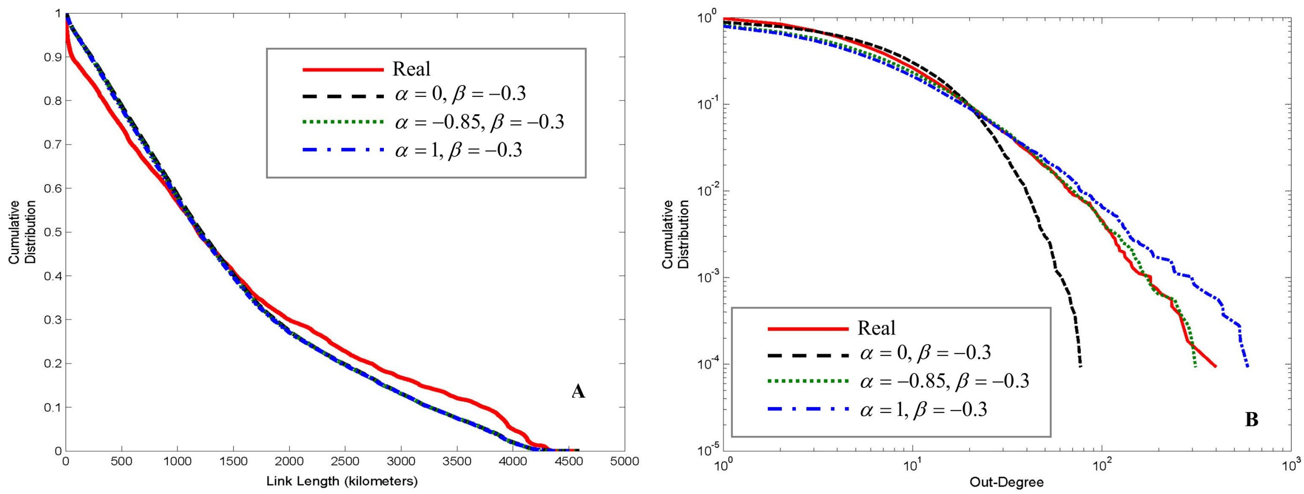


FIG. 8: The effect of the preferential attachment bias on the link length **(A)** and out-degree **(B)** distributions. In the simulations, we have held the "friction of distance" parameter β at its optimal value -0.3, firm location is determined by the population density distribution, and we have varied the "preferential attachment" parameter α. Three network competition models are considered: Geographical Distance Bias without Preferential Attachment (black dashed line, $\alpha = 0,\ \beta = -0.3$); Competition model (green dotted line, $\alpha = 0.85,\ \beta = -0.3$); and Linear Preferential Attachment with Distance Bias (blue dashed-dotted line, $\alpha = 1,\ \beta = -0.3$). Distributions of the actual competition network are shown as red solid lines.

VI. CONCLUSIONS

We have analyzed a large inter-organizational network where the nodes are firms located in the U.S. and directed links represent competition by the nodes forming the link. We focused first on topological properties, and have shown that the competition network exhibits a noticeable asymmetry between the exponentially truncated power law distribution of outgoing competition links and the exponential law governing the in-degree distribution. This characteristic, which is consistent with results of other complex networks [5, 48], can be explained as follows: Firms are not regarded as passive economic entities, but the actions taken by firms could also be seen as determined by and affecting the behavior of other competitors. The exponential law governing the in-degree distribution could indicate a limitation on the firm's capacity to compete with (and thus be affected by) many firms, while the power law governing the out-degree distribution could reflect the ability of competition

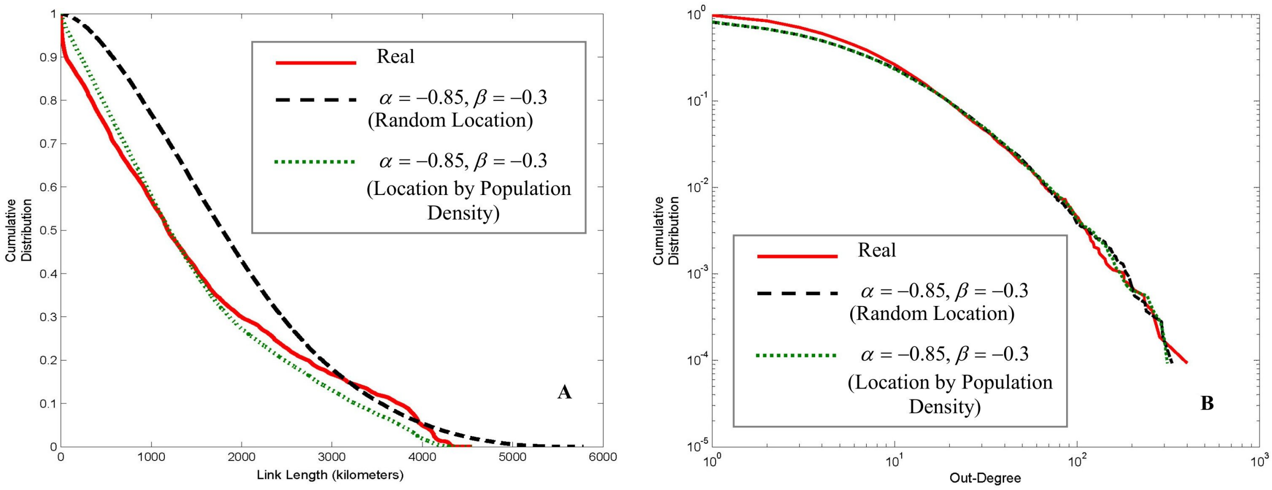


FIG. 9: The effect of firm location decisions on the link length **(A)** and out-degree **(B)** distributions. In the simulations, we have held the "preferential attachment" parameter α and "friction of distance" parameter β at their optimal values 0.85 and -0.3, respectively. Firm location is determined by two distinct mechanisms: Random Location (black dashed line) by which firms are distributed randomly and uniformly on the surface of the contiguous U.S.; and Location by Population Density (green dotted line) by which firms are distributed on the surface of the contiguous U.S. by sampling from the population density distribution. Distributions of the actual competition network are shown as red solid lines.

networks to minimize the effects caused by major events or changes that require significant adjustment in firm behavior. Indeed, the power-law behavior of the out-degree distribution implies that there are only a few firms with many outgoing competition links (*i.e.,* affecting many others), which means that most of the time the competition network will display a low sensitivity to network perturbations. Altogether these results suggest that the structure of competition networks tend to stabilize the dynamics of competition.

The geographical aspect of the competition network has been analyzed in three ways. First, we have shown that the spatial distribution of companies is strongly correlated with the population distribution. This finding emphasizes the important role of environmental and exogenous mechanisms as context for network formation. Second, we have shown that geographic proximity increases the probability of competition following a power law, characterized by a scaling coefficient ("friction of distance") which is considerably lower than values used in other gravity-based models. This result could be explained, for instance, as a consequence of improvements in transport efficiency or communications technology, both of which tend to reduce the value of the friction of distance β. Third, we have analyzed the physical distance between firms, and have shown that the link length probability distribution is well fitted by a slowly decaying exponential distribution with many very short links of length less or equal to 100 km, and extended link lengths of up to 4000 km. This suggests the tendency of competition networks to agglomerate into geographic concentrations ("clusters") of interconnected firms with characteristic size of about 100 km, and with competition links of varying lengths between separate clusters.

Motivated by the above empirical observations, we have proposed a model for the evolution of competition networks, building on recent studies of geographical scale-free networks [14, 25–27]. By comparing simulation results with the empirical observations of the competition network, we have demonstrated that the model is able to reproduce the above features. The competition network model includes two important features that provide insights into the factors governing the origin of competition networks: (1) spatial locations of firms, which is positively correlated with the population density; and (2) stochastic incremental growth governed by nonlinear preferential attachment rule modulated by geographic distance. The stochastic nature of the model suggests that competition networks evolve following cumulative mechanisms where historical accidents, serendipity, and non-linearities induce path-dependent network trajectories.

The model and results presented here are a step towards a coherent theory of interfirm competition network evolution in particular, and evolutionary perspective of economic geography in general. More research is needed in several directions. In this paper we consider the entry of new firms and their links as the only process affecting the size of the network. However, a more realistic description of the evolution of competition networks should take into account the effect of various local events on the large topology of the network, including the formation of new competition links between existing firms, dissolution of existing competition links, shifting (or rewiring) of existing competition links, exit of existing firms, and merging of existing firms. The relative frequency of these local processes will determine (combined with geographic and population density effects) to a large degree the structure of competition networks. Moreover, other tie formation mechanisms operating at the micro-level could be considered such as similarity/dissimilarity in size, performance, or financial indicators between pairs of potential competing firms. While it is theoretically possible to incorporate the above modifications, the scarcity (at this point) of longitudinal competition and firm-specific data over significant time periods make it difficult to validate the model or parameterize it for simulation and prediction purposes.

The simple competition model introduced here offers an evolutionary perspective on economic geography and market structure that significantly extends traditional notions of economic competition and geographical clusters. Combined with proper models of competition dynamics, it also opens up a new range of experimental and analytic possibilities in realistically examining the effect of interfirm competition on firm performance, strategy dynamics, price and output changes, technology diffusion, the emergence of fast-growing geographic clusters (hot spots), and many other phenomena of industry dynamics (*e.g.,* "Red Queen" dynamics [56]). Finally, the model provides a framework to study the ability of competition networks to be resilient (robust) to firm and economic fluctuations.

---

[1] Albert, R. and A.-L. Barabási. 2002. *Rev. Mod. Phys.* 74 47-97.

[2] Amaral, L. A. N., A. Scala, M. Barthélémy, and H. E. Stanley. 2000. *Proc. Nat. Ac. Sci USA* 97 11149-11152.

[3] Boccaletti, S., V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang. 2006 Complex networks: structure and dynamics. *Physics Reports,* 424 175308.

[4] Costa, L. da F., F. A. Rodrigues, G. Travieso and P. R. Villas Boas. 2007. *Advances in Physics,* 56(1) 167242.

[5] Braha D. and Bar-Yam Y. 2007. The statistical mechanics of complex product development: empirical and analytical results. *Management Science* 53(7) 1127-1145.

[6] Amaral, L. A. N. and B. Uzzi. 2007. Complex systems–a new paradigm for the integrative study of management, physical, and technological systems. *Management Science* 53(7) 1033-1035.

[7] Lee, E., J. Lee, and J. Lee. 2006. Reconsideration of the winner-take-all hypothesis: complex networks and local bias. *Management Science* 52(12) 1838-1848.

[8] Barabási, A.-L. and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286 509-512.

[9] Albert, R., H. Jeong, and A.-L. Barabási. 2000. Error and attack tolerance of complex networks. *Nature* 406 378382.

[10] Cohen, R., K. Erez, D. ben-Avraham, and S. Havlin. 2000. *Phys. Rev. Lett.* 85 4626.

[11] Pastor-Satorras, R. and A. Vespignani. 2002. Epidemic dynamics infinite size scale-free networks. *Physical Review E* 65(3) 035108035111.

[12] Hayashi, Y. 2006. A review of recent studies of geographical scale-free networks. *IPSJ Trans. Special Issue on Network Ecology* 47(3) 776-785.

[13] Waxman, B. 1988. Routing of multipoint connections. *IEEE J. Selec. Areas Commun* 6 1617-1622.

[14] Yook, S.-H., H. Jeong, and A.-L. 2002. Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences* 99(21) 1338213386.

[15] Lakhina, A., , J. W. Byers,, M. Crovella, and I. Matta. 2002. On the geographic location of internet resources. *Proceedings of the ACM SIGCOMM Internet Measurement Workshop, Marseilles.*

[16] Gastner, M. T. and M. E. J. Newman. 2006. The spatial structure of networks. *The European Physical Journal B* 49 247-252.

[17] Guimerá, R., S. Mossa, A. Turtschi, and L. A. N. Amaral. 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. U. S. A.* 102 7794-7799.

[18] Porta, S., P. Crucitti, and V. Latora. 2006. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications* 369(2) 853866.

[19] Lim, M., D. Braha, S. Wijesinghe, S. Tucker, and Y. Bar-Yam. 2007. Preferential detachment in broadcast signaling networks: connectivity and cost trade-off. *Europhysics Letters* 79 (5) 58005-1– 58005-6.

[20] Albert, R., I. Albert, and G.L. Nakarado. 2004. Structural vulnerability of the North American power grid. *Physical Review E* 69(2) 25103.

[21] Lambiotte, R., V.D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren. 2008. Geographical dispersal of mobile communication networks. *Physica A* 387 5317-5325.

[22] Eguíluz, V. M., D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. 2005. Scale-free brain functional networks. *Physical Review Letters* 94(1) 18102.

[23] Liben-Nowell, D., J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. 2005. Geographic routing in social networks. *Proc. Natl. Acad. Sci. U. S. A.* 102(33) 11623-11628.

[24] Rozenfeld, A. F., R. Cohen, D. ben-Avraham, and S. Havlin. 2002. Scale-free networks on lattices. *Phys. Rev. Lett.* 89 218701-1–218701-4.

[25] Herrmann, C., M. Barthélemy, and P. Provero. 2003. Connectivity distribution of spatial networks. *Phys. Rev. E* 68 026128-1–026128-6.

[26] Manna, S. S. and P. Sen. 2002. Modulated scale-free network in the euclidean space. *Phys. Rev. E* 66 066114-1–066114-4.

[27] Barthélemy, M. 2003. Crossover from scale-free to spatial networks. *Europhys. Lett.* 63 915-921.

[28] Krugman, P. 1996. *The Self-organizing Economy.* Blackwell, Malden, MA.

[29] Grabher, G. 2006. Trading routes, bypasses, and risky intersections: mapping the travels of 'networks' between economic sociology and economic geography. *Progress in Human Geography* 30(2), 163-189.

[30] Podolny, J. M. 1994. Market uncertainty and the social character of economic exchange. *Administrative Science Quarterly* 39 458-483.

[31] Gulati, R. 1995. Social structure and alliance formation patterns: a longitudinal analysis. *Administrative Science Quarterly* 40 619-652.

[32] Lincoln, J., G. Michael, and A. Christina. 1996. Keiretsu networks and corporate performance in Japan. *American Sociological Review* 61 67-88.

[33] Gulati, R. and M. Gargiulo. 1998. Where do interorganizational networks come from? *American Journal of Sociology* 104(5) 1439-1493.

[34] Stuart, T. E. 1998. Network positions and propensities to collaborate: an investigation of strategic alliance formation in a high-technology industry. *Administrative Science Quarterly* 43 668-98.

[35] Podolny, J. M. 1993. A status-based model of market competition. *American Journal of Sociology* 98 829-872.

[36] Gulati, R. 1999. Network location and learning: the influence of network resources and firm capabilities on alliance formation. *Strategic Management Journal* 20 397-420.

[37] Powell, W. W., D. R. White, K. W. Koput, J. Owen-Smith. 2005. Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. *American Journal of Sociology* 110 1132-1205.

[38] Sorenson, O. and T. E. Stuart. 2001. Syndication networks and the spatial distribution of venture capital investments. *The American Journal of Sociology* 106(6) 1546-1588.

[39] Owen-Smith, J. and W. W. Powell. 2004. Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. *Organization Science* 15 5-21.

[40] Fleming, L. and O. Sorenson. 2001. Technology as a complex adaptive system: evidence from patent data. *Research Policy* 30 1019-1039.

[41] Sorenson, O. 2003. Social networks and industrial geography. *Journal of Evolutionary Economics* 13 513-527.

[42] Google, Inc. `http://www.google.com/intl/en/jobs/locations.html`. Accessed 22 October 2009.

[43] Stouffer, S. A. 1940. Intervening opportunities: a theory relating mobility and distance. *American Sociological Review* 5 845-867.

[44] Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Reading, MA.

[45] Isard, W. 1954. Location theory and trade theory: short-run analysis. *Quarterly Journal of Economics* 68 305- 322.

[46] Zipf, G. K. 1946. The P1P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review* 11 677-686.

[47] Granovetter, M. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology* 91 481-510.

[48] Braha D. and Y. Bar-Yam. 2004. The Topology of large-scale engineering problem-solving networks. *Physical Review E* 69(1) 016113-1016113-7.

[49] Stanley, H. E., L. A. N. Amaral, S. V. Buldyrev, P. Gopikrishnan, V. Plerou, and M. A. Salinger. 2002. Self-organized complexity in economics and finance. *Proc. Nat. Ac. Sci USA* 99 25612565.

[50] Rothenberg, R. B. 1995. Commentary: sampling in social networks. *Connections* 18(1) 104-110.

[51] Lee S. H., P-J Kim, and H. Jeong. 2006. Statistical properties of sampled networks. *Physical Review E* 73 016102-1 – 016102-7.

[52] Solé R. V. and J. Bascompte. 2006. *Self-Organization in Complex Ecosystems.* Princeton University Press, New Jersey.

[53] Hannan M. T. and G. R. Carroll. 1992. *Dynamics of Organizational Populations.* Oxford University Press, New York.

[54] Podolny J. M., T. E. Stuart and M. T. Hannan. 1996. Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991. *The American Journal of Sociology* 102(3) 659-689.

[55] Ruef M. 2000. The Emergence of Organizational Forms: A Community Ecology Approach. *The American Journal of Sociology* 106(3) 658-714.

[56] Barnett W. P. and M. T. Hansen. 1996. The Red Queen in Organizational Evolution. *Strategic Management Journal* 17, 139-157.