Supple Networks: Preferential Attachment by Diversity in Nascent Social Graphs

JAMESON K. M. WATTS and KENNETH W. KOPUT
Eller College of Management, University of Arizona ∗
(*e-mail:* jamesonw@email.arizona.edu)

Abstract

A preference for diversity has been identified as an important predictor of tie formation in certain networks, both social and organizational, that also exhibit a high degree of suppleness–the ability to retain their general form and character under stress (Durkheim, 1893/1997; Powell *et al.*, 1996; Powell *et al.*, 2005; Koput & Gutek, 2010). Extant models of preferential attachment, based on popularity, similarity, and cohesion, meanwhile, produce exceedingly brittle networks (Albert *et al.*, 2000; Callaway *et al.*, 2000; Holme *et al.*, 2002; Shore *et al.*, 2013). A model of preferential attachment based on diversity is introduced and simulated, demonstrating that a preference for diversity can create a structure characterized by suppleness. This occurs because a preference for diversity promotes overlapping and redundant weak ties during the early stages of network formation.

Keywords: *social networks, complex networks, diversity, resilience, preferential attachment*

1 Introduction

Social ties seldom form at random. Attachment of nodes, whether persons or organizations, is, instead, biased; or, in network parlance, preferential. There is a long stream of literature, for example, that emphasizes the role of similarity as a bias, or preference, in tie formation among social actors. The concept of homophily–the principle that likeness attracts, or "birds of a feather flock together," is widely-viewed as the predominant, naturally-occurring mechanism of social tie formation (Wimmer & Lewis, 2010; McPherson *et al.*, 2001).

Resemblance, whether in the form of shared attributes, common interests, or physical proximity, provides a basis for interaction as well as affinity. The social structures that emerge tend to be primitive and clan-like: segmented networks comprised of homogeneous nodes and exhibiting what Durkheim (1893/1997) referred to as "mechanical solidarity." The "tribal" logics (Durkheim, 1893/1997) characteristic of such structures precipitate the

∗ We thank the Office of the Vice President for Research at the University of Arizona for an allocation of time on the High-Performance/High-Throughput Computing (HPC/HTC) cluster. We also thank Ron Breiger, Woody Powell and the participants of the 19th annual graduate workshop on complexity in the social sciences at the Santa Fe institute for helpful comments on early drafts of this manuscript. Mail can be sent to: Department of Marketing, Eller College of Management, The University of Arizona, McClelland Hall 320, P.O. Box 210108, Tucson, AZ 85721-0108. Phone: 303-335-9232.

patterns of preferential attachment in tie formation and hierarchy in degree distributions that are mimicked in many extant models of complex networks (see e.g. Barabási and Albert, 1999; Barabási and Frangos, 2002). Within such exogenously-conformed networks, ties are seen to accumulate around the eldest, most-central and most-cohesively connected as nodes begin to diverge endogenously on such properties. The resulting social bonds are theorized to be dyadic, as homophily relates to the similarity between a focal node and each of its partners taken separately (Durkheim, 1893/1997). Consonantly, in simulation studies of networks based on preferential attachment by popularity, similarity, or cohesion, the resulting structures are exceedingly brittle–the targeted removal of a small number of nodes or edges breaks such networks apart, altering their basic properties, character and function (Albert *et al.*, 2000; Callaway *et al.*, 2000; Holme *et al.*, 2002; Shore *et al.*, 2013).

Diversity, as a basis for tie formation, has often been viewed as a product of intervention or design (see, e.g. Reagans et al, 2004). In this view, the construction of social groups composed of dissimilar members arises as prescriptive for obtaining both individual and collective outcomes as themselves varied as getting a job, assimilating into a new culture, pursuing scientific discovery and technological innovation, shepherding a social movement, and colluding (Granovetter, 1973; Taylor, 1989; Baker & Faulkner, 1993; Hargadon & Sutton, 1997; Powell *et al.*, 1999; Shemtov, 2003; Powell *et al.*, 2005). More recently, empirical studies have begun to identify a preference for diversity as an important, unimposed predictor of tie formation (Powell *et al.*, 2005; Koput & Gutek, 2010).

Difference, in skills, attributes, and interests, provides a basis for complementarity as well as friendship (Durkheim, 1893/1997; Pachucki & Breiger, 2010). Arising from interdependence, rather than likeness, both the social bonds and collective social structures that emerge tend to be stronger than those based on similarity. Solidarity is "organic", rather than mechanical (Durkheim, 1893/1997). We seek out dissimilar others because they complement us, making us feel less incomplete as a result (see e.g. Durkheim, 1893/1997 p. 54). Social bonds are strong because diversity is structural, as the attraction in difference reflects the specialization and division of labor of more advanced societies. This logic readily extends beyond a dyad to a circle of friends, or portfolio of ties: we should prefer not just that each of our partners differs from us, but that they differ from each other as well. Accordingly, in a number of empirical settings featuring a preference for diversity in tie formation, the networks appear remarkably supple–maintaining their basic character and function as key nodes or edges are removed (see e.g. Powell *et al.*, 2005), or even inserted (Koput & Gutek, 2010).

Does a preference for diversity, then, create suppleness? To begin to answer this question, we develop and simulate a model of preferential attachment by diversity. We examine the nascency of the network that results, in comparison to that of referent networks generated by models of either random attachment or preferential attachment by popularity or clustering. In this way, we hope to reveal the development of suppleness. Before proceeding, some remarks are in order: first, on suppleness; then, on diversity.

2 Suppleness

In lay terms, suppleness suggests a certain softness or pliability. Supple objects bend gracefully, and are able to return with ease–but equally able to hold the new position without strain. Supple networks display resilience when key nodes or edges are removed, retaining or restoring their basic properties while maintaining their basic character and function. But suppleness, as we introduce the concept here, extends beyond the notion of resilience. Resilience is anchored and episodic. Resilient networks "snap back" from each induced change to a fixed, prior state, at the same rate and in the same manner (Bishop *et al.*, 2011). Supple networks become more resilient with each instance of change, responding with increasing ease. Further, supple networks can absorb, as well as rebound from, change. Whereas resilience implies elasticity, suppleness implies plasticity. While resilience resides in the structure of a network, suppleness comes from the underlying generative process.

An example may help. Throughout the 1990's, observers repeatedly forecast the consolidation of the biotechnology industry. The industry was comprised of hundreds of small, startup firms who gained access to much-needed resources through strategic alliances with a variety of partners: venture capitalists, universities, legal firms, research labs, hospitals, government agencies, and conglomerate chemical and pharmaceutical firms. The latter dominated the landscape of partners for commercializing the new therapies being discovered by dedicated biotech firms. "Big Pharma" had the resources and incentive to buy up the most prominent and productive biotech firms, which also were the most centrally-connected (Powell *et al.*, 1996). The large pharmaceutical firms "cherry-picked" the most promising biotech firms, effectively removing the most central nodes from the network (Powell *et al.*, 2005). Yet, the inter-organizational network maintained its basic character, properties, and role as the locus of innovation–so much so that, in a number of cases, the once-acquired biotech firms were spun back out as independent firms, reintroduced as nodes in the network with many, or even most, of their ties restored as edges. The industry did not consolidate, despite the number of targeted "buy outs." The inter-organizational network in biotech exhibited a suppleness that goes beyond resilience. In subsequent research covering the same time period, a preference for diversity in tie formation was revealed as a key part of the logic of attachment in the biotech industry (Powell *et al.*, 2005).

3 Diversity

Diversity, as it suits our purpose, is attributional: meaning that it stems from exogenous characteristics of nodes–size, history, demography, status, financing, wealth, products, vocation, technology, skill and so forth. Extant models of preferential attachment tend to treat nodes as in-situ homogenous. Nodes become different only in endogenously-determined egocentric properties, such as numbers of ties. To the extent that these models of tie formation are to describe social or organizational networks, the lack of exogenous variety in nodes is limiting. When all birds are of the same feather, homophily's adage can no longer explain why some flock together while others remain apart.

Attributional diversity is generally constructed of two or more dimensions, the first of which is nearly always considered to be variety. That is, node attributes must vary in some way–there must be different kinds of nodes at a demographic level, or different sorts of subjects or events with which nodes affiliate. The preference for diversity, then, is a bias for forming new attachments to partners that vary, in some way, from partners to which a node is currently attached.

Once variety is established, two common additions to the conceptualization of diversity are balance and disparity.[1] Balance is a function of the uniformity, or evenness, of the distribution of varietal types. For instance, in the simplest case where the attribute characterizing nodes has two categorical levels, variety is established once a node is attached to any partners of each type. Consider a biotech firm with six partners characterized dichotomously as universities or pharmaceutical companies. Suppose the node has five university partners and one pharmaceutical partner. The variety is two. Now suppose it has three of each form–the variety is still two. Yet, it is hard to consider the alternative tie portfolios as equally diverse. The former is the least balanced it could be, while the latter is fully balanced–and certainly seems more diverse as a result.

Disparity, as a third dimension of diversity, can only be considered when the nodes are characterized by an attribute that has a numerical level of measurement. While disparity is of widespread interest as an outcome of social processes, it is less often incorporated into the discussion of diversity as an antecedent to structure, where consideration of qualitative, categorical attributes (e.g. gender, race, religion, socioeconomic status, and the like) is dominant.

4 Archetyping Attachment

We noted at the outset that social ties seldom occur at random. Yet, equally rarely are they entirely predetermined, whether by similarity, diversity, or some other formative basis. Chance encounters provide opportunities for recognizing commonalities or complementarities. Moreover, real actors have many attributes on which tie formation may be based. When the attributes which form the basis of attachment vary among dyads, the resulting process might well seem random. If so, a random network might well portray the aggregate properties of many real networtks. Even though not formalizing the underlying causal mechanisms, a random model of attachment might well serve as a useful baseline for us because they produce networks that are much less brittle than other established formulations. Erdős and Rényi (1960) are responsible for deriving the most prominent example of a random network, which has come to be known as the Erdős-Rényi (ER) model of random attachment.

Barabási and Albert (1999) extended the ER model to non-random rules of attachment. In particular, the Barabási-Albert (BA) model of preferential attachment introduces a bias to tie formation based on popularity, an endogenous property. In this model, new ties go to nodes that have the most previously established ties, or total partners (Barabási & Albert, 1999). The obvious tendency for all ties to concentrate on the two nodes that, by whatever logic, are first tied, is typically attenuated through aggregated rounds with selective tie assignment. Rather than being defined by the allocation of a single tie, each round encompasses the formation of several ties, only one of which can be received by any given node in a single round with as many nodes receiving as there are new ties. While the preference for popularity of the BA model is certainly the most studied, rules for

[1] Various terms have been used by researchers to describe the three basic dimensions of diversity (Pachucki & Breiger, 2010; Blau, 1977; Herfindahl, 1950). However, each is qualitatively the same and analogous to those used in the current work.

attachment can be based on other endogenous node-level characteristics besides number of ties, such as duration of exposure, cohesion, or geodesic distance.

Both random and rule-based attachment can vary in the way that the probabilities of receiving a tie are allocated. In the original ER model, ties are randomly formed with equal probability between nodes in a graph of fixed initial size (Erdős & Rényi, 1960). The probability of receiving a tie is thus uniformly distributed across a population of nodes and held constant. The difficulty with this model is that most real world networks are dynamic, such that new nodes are added (or removed) over time. Barabási and Albert (1999) updated the ER model so that the attachment process is initiated by the appearance of new nodes. This revised ER model evolves to a steady state where the probability of any degree $P(k) \sim exp(-\beta k)$. Such a revision serves to facilitate comparisons across time for graphs evolving under distinct regimes, while preserving the basic character of the original ER model.

In both the ER and the BA model, then, the generating function for each regime proceeds in the following manner. Upon initiation, m_0 nodes are created. In each subsequent step, one new node is introduced and given m ties to existing nodes provided a tie does not already exist. This process repeats t times such that the resultant network contains $m_0 + t$ nodes and $m * t$ total ties. Under the ER regime, existing nodes are equally likely to receive a new tie provided one doesn't already exist. However, under the BA regime, nodes with many existing (previously received) ties are more likely than others to receive a new tie.

The power of this type of generating function lies in its ability to hold network density constant such that macro-level characteristics can be reliably averaged across instantiations and compared across models. Indeed, the ER model evolves to a state in which the number of connections per node (degree) is exponentially distributed whereas the BA model evolves to a scale-free distribution. Yet, comparisons between the two models can be reliably made because both the number of nodes and the number of ties are held constant at arbitrary points in time–it is merely the configuration of ties which differ.

Most algorithms for complex network growth can be understood as a biased version of the generating function that produces the ER model (see e.g. Skvoretz *et al.*, 2004). Indeed the BA model introduces one such bias, namely popularity. Whereas ER ties are formed based on draws from a uniform distribution, BA ties are formed based on draws from a distribution defined by structural characteristics of the existing nodes. Thus, one potent strategy for creating new models is to redefine the attachment distribution at each time step. This strategy is attractive for its elegance–distributions are created for an arbitrary number of features, these distributions are summed, and choices are based on draws from the combined distribution. However, this introduces the unpleasant problem of defining *a priori*, how to weight the contribution of each feature. Different weights can produce vastly different outcomes.

A more tractable approach is to assume that attachment mechanisms are nested. In other words, a subset of the available nodes is identified using one criterion and then the final choice is based on weighting nodes in the subset by a second criterion. This is the tactic used by Holme and Kim (2002) in their extension of the BA model to include tunable clustering. Moreover, this approach finds widespread support in work on the psychology of judgment and decision-making (Boland *et al.*, 2012), and in many economic models of choice (Tellis, 1988). Lastly, even if one finds it difficult to determine *a priori*, which attachment mechanism is used first, inductive exploration is limited to $c!$ versions of the model, where c is the total number of criteria. While it should be noted that our results are qualitatively the same when attachment probabilities are combined rather than nested, we present the latter as it conforms more readily to the heuristic nature of human decision-making and results in a more elegant and intuitive generating function. Specifically, we offer a model based on a preference for diversity as the second step in a two-stage, nested attachment process.

5 Preferring Diversity

Following from our discussion in Section 3, any formalization of a diversity preference must (at a minimum) include mechanisms that address both variety and balance. Incentives for variety are evident in the Durkheimian notion of complementarity–ties to synergistic (but distinct) partners have a value greater than that achieved by an equal number of ties to partners of the same type (Pachucki & Breiger, 2010). Granted, the level of complementarity may vary according to the composition of types in an actor's ego network. For instance, firms might value the synergy of a bank and accounting firm more than they would the synergy of a bank and advertising agency. Yet, organic solidarity is still achieved when each tie is to a distinct partner rather than to partners of the same type (i.e. two banks). Admittedly, variance in the level of synergy by dyad is of practical concern–one would expect the strength of solidarity to vary accordingly. While acknowledging the substance of such arguments, we can assume for now that a variety dimension of diversity is addressed if we account for *equivalent* synergies by dyad across the set of defined node types. We leave a more nuanced account of varying complementarity to future research.

While balance is in and of itself an integral dimension of diversity, a more subtle (and structural) mechanism can also prompt balance-seeking behavior, albeit indirectly. This occurs when an actor is concerned with varietal redundancy. This dimension of diversity is absent in the Durkheimian account, but implied in studies that deal with the evolution of networks over time. As a social phenomenon, dynamism in ego networks is commonplace. Friends drift apart and relatives perish, just as new relationships are formed. In organizational networks, firms are acquired or go bankrupt, just as new alliances are established. Accordingly, solidarity (and the requisite variety) must be managed in light of the uncertainty introduced by these random events. Thus, *depth* as a concomitant dimension of diversity is motivated by considerations of social dynamics.

When depth is accomplished, actors can use a subset of complementary ties, while redundant ties remain dormant. Under this description, the benefits of additional depth clearly diminish in the amount of current depth–a tie to a second bank is more valuable than a tie to a third bank and so forth. Yet, an assumption of diminishing returns to redundancy naturally leads to balance. Actors prefer additional depth for types where they have little to begin with–a weighting amongst positive valuations–which, given enough time, inevitably leads to parity (or balance) across types in an ego network.

Balance as an outcome of varietal redundancy will only emerge if the value of depth is equivalent by node type. If certain types are introduced or removed from a network with higher (lower) frequency, then the value of depth should reflect these probabilities. However, maintaining balance across node types is a reasonable (often optimal) heuristic for achieving depth when information about the true probabilities is either sparse or costly to attain (Stinchcombe, 1965; March, 1991). For instance, knowledge about the true distribution of new and terminated partner types is imprecise during the early stages of network formation (i.e. nascent social graphs). Nevertheless, actors may be more or less strategic in how they approach the problem of depth as information becomes more precise over time. If for instance, a firm learns that certain partner types are more likely than others to be acquired by competing firms, they might decide to maintain a higher proportion of these partners in their network. However, an explicit learning component is beyond the purview of the current work. A model of diversity preference is sufficient if it addresses the heuristic strategy–a uniform distribution of depth value across node types.

Lastly, any model of diversity preference should allow for benefits to level off as diversity is achieved. Granted, certain empirical research has shown that diversity actually becomes a liability (turns negative) after crossing some threshold; however, this is typical in mature networks where concerns over legitimacy conflict with the solidarity argument (Owen-Smith *et al.*, 2014). Our focus here is on the forces which describe nascent social graphs. Moreover, many empirical studies that incorporate attributional diversity in their models present data in a range with diminishing (rather than negative) returns. This is despite significant negative coefficients on the squared terms (see e.g. Cohen & Broschak, 2013) that would otherwise indicate a peaking benefit. While, the liabilities of inordinate diversity have theoretical merit, for now we focus on a benefit that increases at a decreasing rate.

Naturally, if the number of total possible attributes is fixed, then once an ego network has nodes of all possible types, variety cannot be increased and the value of variety is fully realized. However, the value of *maintaining* variety in the presence of uncertainty will still drive behavior. This happens via increases in depth and the process of rebalancing. As previously noted, the value of depth decreases in the amount of current depth. To the extent that balance is simply the outcome of increasing depth by a heuristic, its value should also diminish as depth is achieved. However, the value of balance is more than just realized depth. Rather, balance is a primary motivation in diversity and should *negatively* weight additions to an ego network that place it *out* of balance. This is not to say that additions which adversely affect balance should be negative overall–varietal redundancy is still desired. Rather, redundant ties that place a network out of balance, should be valued lower than those which restore balance.

In sum, a model of diversity preference should 1) highly value any changes that increase variety, 2) value changes that increase depth, 3) value increases in depth that improve balance more than those that do not, and 4) diminish as variety, balance and depth are achieved. In the next section, we provide a formal model of the diversity preference that addresses these four properties.

5.1 A Model for Diversity Preference

In our conception of diversity, nodes vary along an arbitrary, non-structural category in which d captures the number of mutually exclusive classifications. Under these circumstances, variety and balance are maximized when the probability of selecting two nodes (at random) of the same type is lowest (Gibbs & Martin, 1962; Herfindahl, 1950). Formally, given some neighborhood, the probability of selecting a node of type d_j, is equal to the proportion of nodes of that type in the neighborhood, which we denote p_j. We can then specify a diversity index as one minus the probability of selecting a node of the same type twice.

$$H = 1 - \sum_{j=1}^{d} p_j^2 \tag{1}$$

However, a *preference* for diversity implies forecasting. Actors possess a current ego network and weight the contribution of a hypothetical new node based on type. We call this hypothetical contribution the diversity margin δ_{ij}, of adding a node with type d_j to the neighborhood of $node_i$. Values are determined by the diversity index of $node_i$ after the addition of a hypothetical new node of type d_j, and subtracting out the original diversity index,

$$\delta_{ij} = H_{ij} - H_{i0} \tag{2}$$

such that δ_{ij} is the diversity margin for $node_i$ after the addition of a hypothetical node of type d_j.

Since H can vary between 0 and $(d-1)/d$, δ will take on values between $(1-d)/d$ and $(d-1)/d$, which violates the requirement that depth-increasing changes have a positive value. To illustrate, imagine $d = 2$ such that any given network neighborhood is comprised of nodes of type A and nodes of type B. Further consider the specific neighborhood $\{A, B\}$ with diversity index $1/2$. Clearly, the addition of a new node of type A or B will reduce the index in this neighborhood to $4/9$ such that $\delta = -1/18$.

One way to circumvent this issue is to supply the diversity margin as a parameter to the exponential function. This has the desired effect of maintaining a positive value for depth-enhancing additions $(\delta_{ij} < 0)$, while emphasizing depth that restores balance $(\delta_{ij} > 0)$. Moreover, Equation (1) implies that $e^{\delta_{ij}}$ naturally diminishes as variety, balance and depth are achieved for $node_i$. Thus, we can specify the preference for diversity as a matrix of weights Ω , where ω_{ij}, represents the preference of $node_i$ for connections with new nodes of type d_j.

$$\omega_{ij} = e^{\delta_{ij}} \tag{3}$$

5.2 Generating Function

We use both the ER and the BA model as a starting point for our algorithm and provide simulations that incorporate a preference for diversity under each regime. We also implement the Holme-Kim (HK) model of tunable clustering as an archetype of cohesion-biased attachment. Their model provides a conceptually relevant contrast to diversity-seeking behavior and uses a similar 2-stage selection process. The specifics of our implementation proceed as follows.

Initiate the graph with $m_0 > m$ new nodes. In each subsequent step, add one new node to the graph and randomly assign this node a classification of d_j. Then create m ties between the new node and existing nodes using the following process for each attachment.

From the set of all existing nodes, select m at random (or weighted by popularity under the BA regime) provided a tie does not already exist. From this set, make a random choice weighted by the potential benefit of adding a node of type d_j to the existing node's neighborhood. Our pseudo-code is presented below followed by a discussion of macro-level features of the model along side implementations of ER, BA and HK with the same number of nodes and ties.

```
Input:
- m: the number of initial nodes
- n: desired number of nodes in the final graph
- d: number of distinct node types
- regime: a switch for first-stage node selection (ER or BA)

create m initial nodes
for i ← m+1 to n do
    create a new node_i
    randomly classify node_i as type d_j
    generate Ω  according to equation (3)
    for c ← 1 to m do
        if regime is ER then
            choices = m random draws from existing nodes without replacement
        end
        if regime is BA then
            choices = m random draws from existing nodes, weighted by degree,
            without replacement
        end
        form a tie between node_i and a random draw from choices weighted by ω_j
    end
end
```

Algorithm 1: Preference for Diversity under ER or BA

6 Network Descriptives

Figure 1 shows degree distributions of graphs with 1,000 nodes on a log scale (top) and the evolution of a graph's variance in degree from m to 100 (bottom). Graphs are arranged with ER models on the left and BA models on the right. A baseline model corresponding to the parent regime is shown alongside implementations of a clustering and of a diversity bias to the baseline model. As you can see from the degree distributions under the BA regime, the scale-free nature of accumulated advantage appears to hold even when a preference for diversity is added to the model (BA+D). This feature is detected by examining the linear relationship between the log of degree k, and the log of its frequency in a given network. Specifically, scale free graphs are identified when the probability of a given degree $P(k) \sim k^{-\lambda}$, where λ is often found to be between 2.0 and 4.0 in real-world networks (Barabási & Albert, 1999; Clauset *et al.*, 2009).

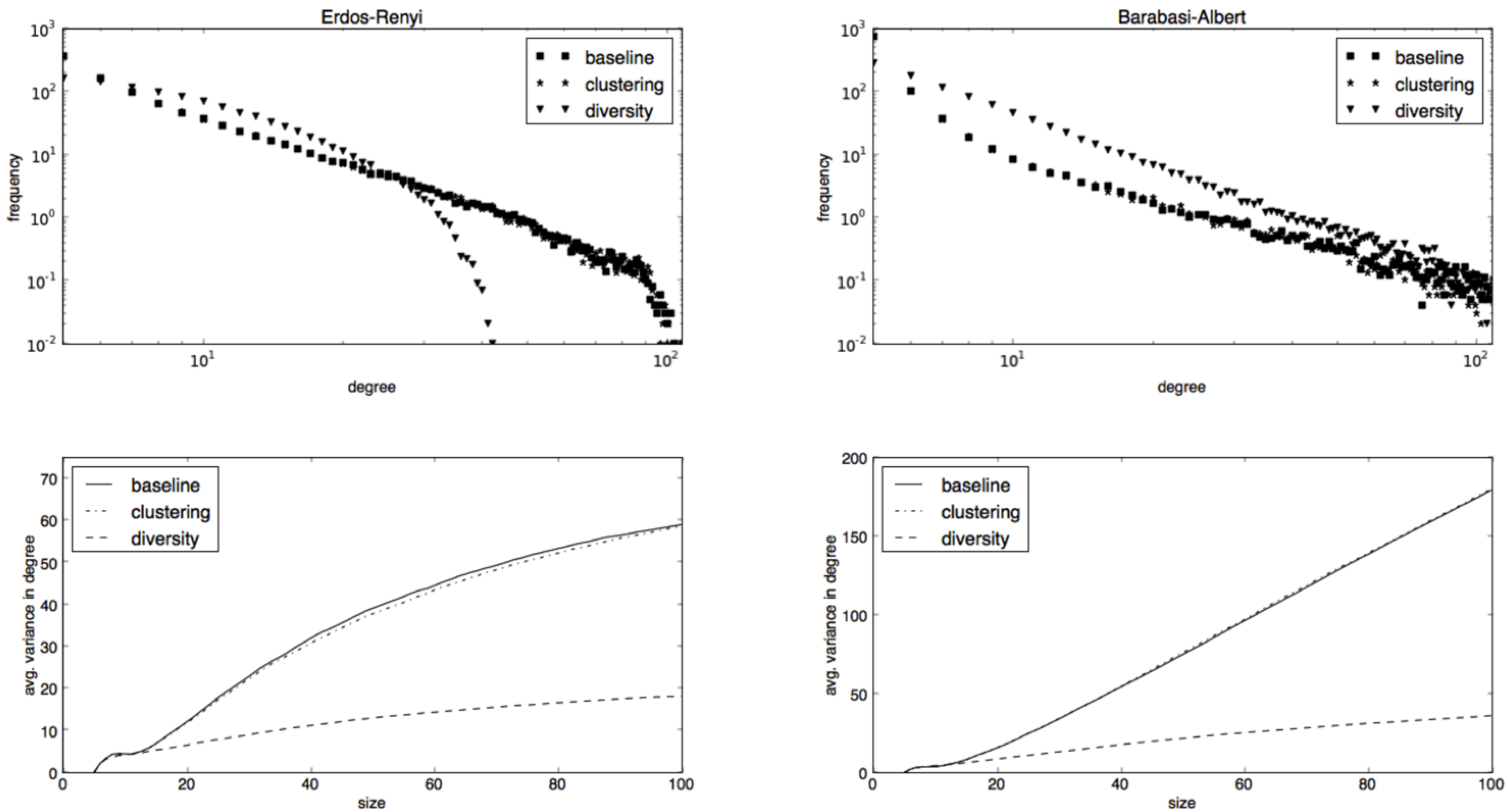

Fig. 1. Degree distributions for graphs with n=1,000 (top) and average variance in degree as graphs grow from n=5 to 100 (bottom row). All values are averaged across 100 random trials.

For a pure BA model, λ evolves to approximately 2.9; however, this is expected only after the network has grown quite large and may exist only in the tail of the distribution. Indeed, Figure 1 suggests that the baseline BA model has yet to exhibit a truly scale-free distribution. The simulated network consists of only 1,000 nodes, so this is expected. However, the graph trends towards scale-free much quicker with the addition of a preference for diversity to the standard BA model. With 1,000 nodes averaged over 100 instantiations, and $m_0 = m = 5$, our generating algorithm for BA+D puts λ at approx. 3.05. This fact is noteworthy in its own right; however, more precise methods for estimating λ involve larger simulations and tests for goodness of fit at both low and high k (Clauset *et al.*, 2009). Thus, we leave a complete analysis of scaling to future research and instead focus on macro-level characteristics more closely related to suppleness. However, it should be reiterated that our preliminary analysis suggests a preference for diversity is not inconsistent with the the scale-free property in graphs with $n \sim 1,000$. This is an important consideration given that several recent studies have identified scale free distributions in smaller networks ($n < 2,000$). Indeed Powell *et al.*, (2005) study such a network that is both scale free and supple, and explicitly reject the BA process as a primary driver of attachment.

As you can see from the two graphs at the bottom of Figure 1, variance in degree is generally much higher under the BA regime. This is not surprising since preferential attachment based on popularity promotes a scenario in which the rich-get-richer and the poor remain as such. However, variance is increasing monotonically under both the ER and the BA regime. This is also understandable, since older nodes will gain more ties by virtue of their presence during many random draws in the history of the graph. In all cases, the addition of a preference for diversity reduces variance in degree. This is because the effect of a node's age (ER) and the popularity bias (BA) are attenuated by a preference for diversity–especially early in the growth process.

Equation (2) implies that the potential benefit of adding a new node is decreasing in the size of one's network neighborhood. Take for example the set $\{A,A,B\}$ and the set $\{A,A,A,B,B\}$. Clearly the addition of a B-type node will have greater benefit to the first set than it will to the second. This means that existing nodes with smaller neighborhoods will attract connections from new nodes at a higher rate than those with larger neighborhoods. This is especially true during the early stages of network growth when there are few nodes overall. In other words, the nodes available during stage one of the selection process are smaller when the network is young. Thus existing nodes with small neighborhoods are more likely to be considered and more likely to exert their preference for diversity.

The attenuating effect of a preference for diversity is shown in Figure 2. Essentially, as existing nodes with small neighborhoods fight for attachment with the new nodes coming online, connections are more evenly spread throughout the graph, creating something like a crystalline lattice. The graphs in Figure 2 show networks of size $n = 10$, $n = 30$ and $n = 50$ under pure BA (top) and BA+D (bottom). While connections are severely concentrated under the pure BA model, they are more spread out when a preference for diversity is added to the growth model.

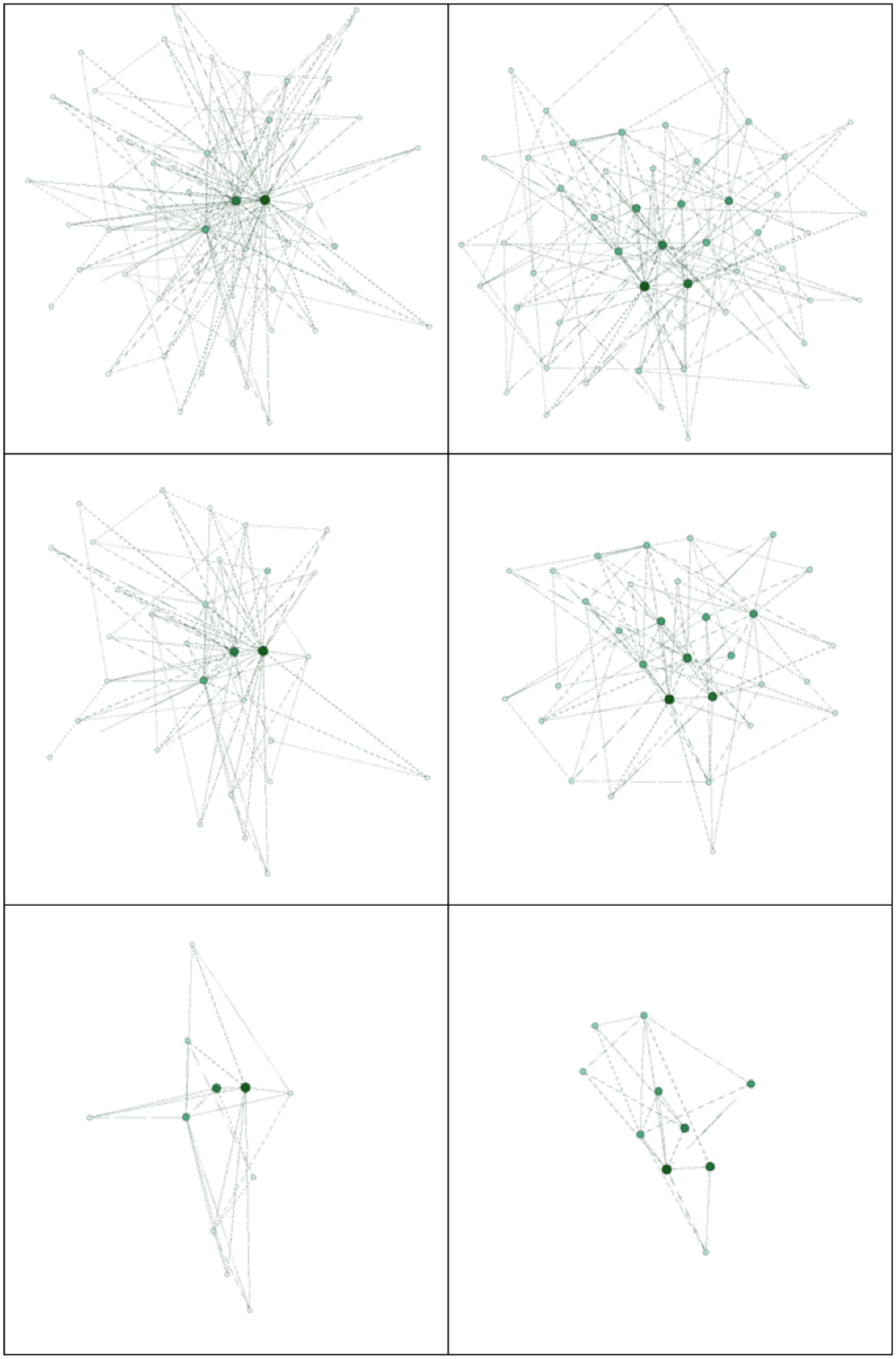

Fig. 2. Patterns of growth for BA (top row) versus BA with a Diversity Preference (bottom row).

7 Results

Suppleness, as we conceive it, is the ability to retain structural integrity while undergoing change. Supple networks are able to maintain or restore global properties when subject to the loss or removal of nodes or edges. Here, we focus on how each network fares over time as the most central nodes are systematically removed. While the literature on network fragility considers, variously, the removal of nodes and edges (Callaway *et al.*, 2000), we focus solely on the removal of nodes, and specifically those with the highest degree-based centrality, for two reasons. First, edge removal is subordinate–the removal of a central node eliminates many edges. Second, node removal seems of more paramount concern when translating to real-world scenarios, particularly as an analog to mergers, acquisitions and exit in industrial networks such as that of the biotech industry (Powell *et al.*, 1996; Powell *et al.*, 2005) or drop-outs, fallbacks, or commencements in classmate networks such as the one studied by Koput and Gutek (2010).

7.1 Post-hoc Attacks

Existing research generally focuses on post-hoc attack strategies (e.g. Holme, 2002) in which nodes are strategically pruned after a network grows to a certain size. This is useful because it can reveal the structure upon which metrics like average path length are based (Girvan & Newman, 2002). More importantly, it can reveal when those structures begin to fail (Callaway *et al.*, 2000). The analysis presented here may be used as a comparison to results found in existing studies on the strength and vulnerability of generated networks. In each case, degree-based centrality scores are calculated and nodes with the highest scores are removed in order of highest to lowest degree. In other words, those nodes with the most connections are removed first. The scores defining which node is to be removed may be calculated once at the beginning, or recalculated after each node is removed (Holme *et al.*, 2002). In the figures below, we use the former strategy, but note that our results are qualitatively the same regardless of the specific implementation.

Figure 3 shows the results of a degree-based attack on a graph's inverse geodesic and the size of the largest component. These are two of the most commonly used metrics when studying a network's response to attack (Callaway *et al.*, 2000; Holme *et al.*, 2002) [[–*should we continue to use the word "attack" or should we use something else?*–]]. The inverse geodesic of each graph is a convenient interpretation of average path length (APL) in which the calculation for fragmented graphs can be handled gracefully. APL is calculated as the average of all shortest paths between nodes in a graph; however, in a fragmented graph there are several paths that do not exist and so are considered of infinite length–a value, which affects the average to say the least. If instead, we use the inverse of the distance, 1/infinity can be defined as zero and the calculation proceeds benevolently.

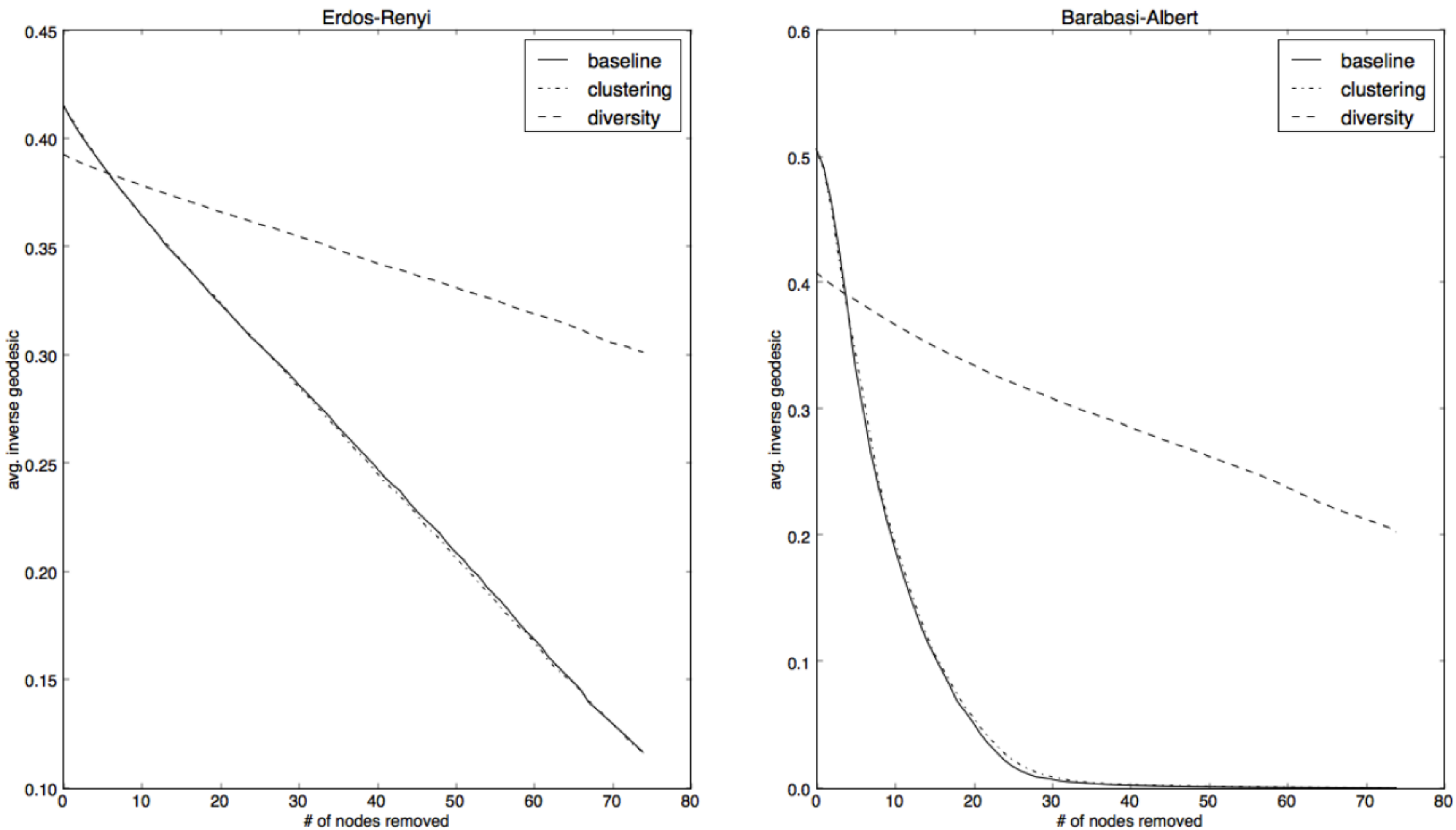

Fig. 3. Post-hoc attack after network is grown to n = 300. Attack is based on degree-centrality. Values averaged over 10 random trials.

In essence, each metric is capturing the level of overall connectedness in the network. If large chunks of the graph become disconnected, the size of the largest connected component will drop dramatically. If large chunks of the graph become isolated–either completely cut off or sparsely connected–the inverse geodesic drops dramatically. This is because nodes either can't reach each other, or their shortest paths all go through one or two edges making the overall trip much longer. You can see from each figure that the addition of a preference for diversity to either the ER or the BA model significantly reduces a graph's vulnerability to this sort of attack.

7.2 Concurrent Attacks

While existing research generally focuses on post-hoc attack strategies (Holme *et al.*, 2002), we suggest that these after-the-fact approaches often address questions of vulnerability rather than suppleness. This is because the process essentially measures the number of attacks a network can withstand before a collapse. We've shown that a preference for diversity can reduce a network's vulnerability, and yet, this is only one component of suppleness. Supple graphs must be able to retain their character while undergoing change. This means that whichever mechanisms are responsible for new connections must also respond dynamically to exogenous change. Thus, we implement a process of strategic attack that coincides with growth.

In this analysis, an attack consists of finding and removing the s, most connected nodes in the network provided some frequency threshold a, is met in the current time period and a grace period g, has passed. Figure 4 shows the evolution of our generated networks with $s = 1$, $a = 50\%$, and $g = 30$. Results are similar using a wide range of parameterizations. Specifically, our algorithm for the concurrent attack strategy is implemented inside the main loop of our generating function in the following manner. Prior to the creation of a new node, execute an attack of size s, if a randomly generated value between 0 and 1 is less than or equal a, and the size of the graph is greater than g.

As you can see from Figure 4, networks biased with a preference for diversity tend to retain their global characteristics while the others undergo qualitative change. This is especially noteworthy under the ER regime in which a bias towards diversity responds better than the baseline (random) simulation. Random graphs are supple by their very nature and thus provide a conservative model against which to compare a preference for diversity. The results under the BA regime are more drastic, but also more expected. When a network is biased towards popularity, many shortest paths will flow through nodes with the highest degree. It follows that removal of high-degree nodes during attack will have an outsized impact on the network's characteristic path length. However, adding the preference for diversity as part of the selection criteria adds a level of protection to networks, which would otherwise evolve to an exceedingly brittle state (Albert *et al.*, 2000)

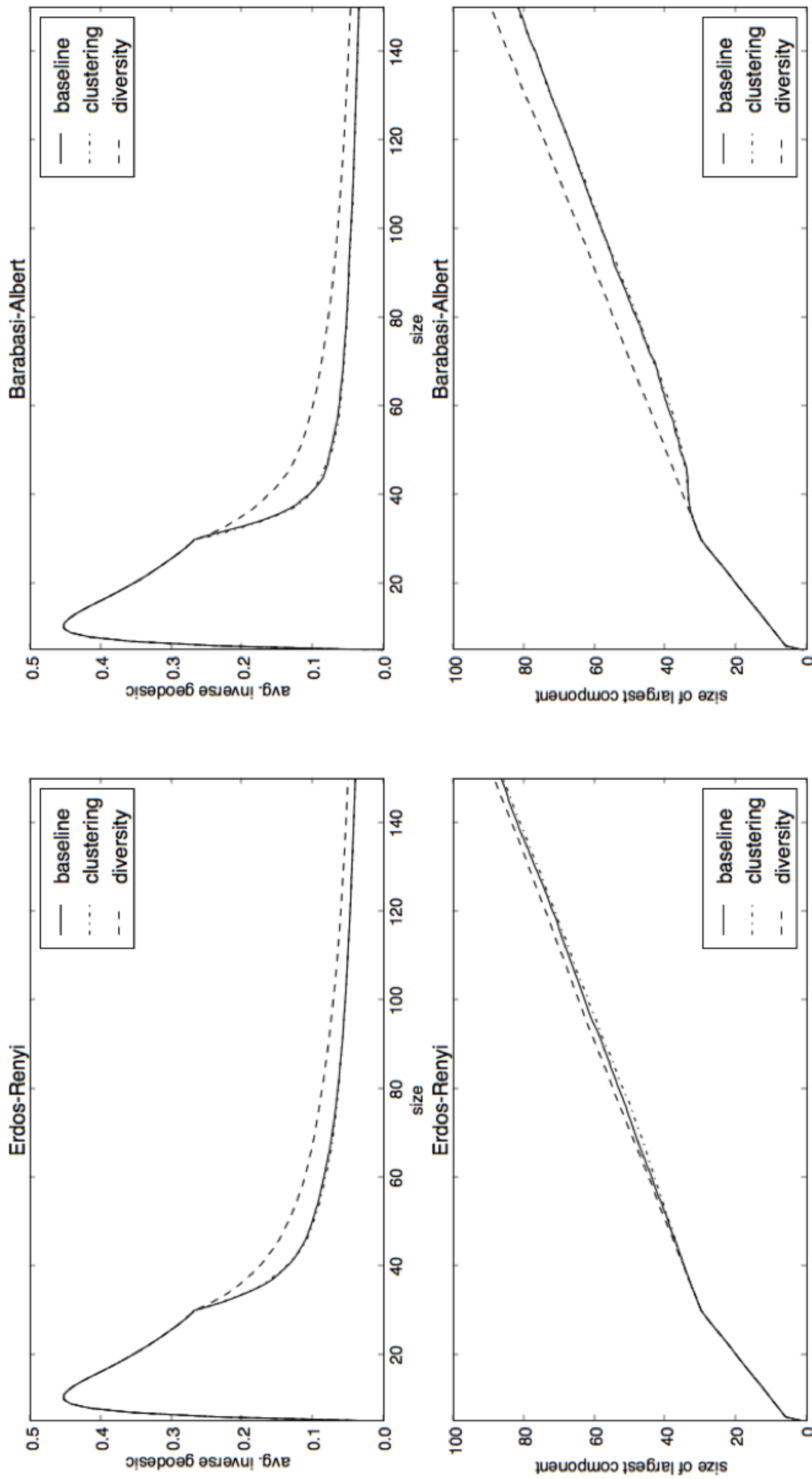

Fig. 4. Average path length and largest component with attacks starting at n=30. All values are averaged across 100 random trials.

While the inverse geodesic and size of the largest component are the most frequently used measures in the literature on network vulnerability, they do not capture all of the structural properties affected by exogenous attack. Nor do they reveal the way in which change is incorporated. Recall from our definition that supple networks exhibit both some level of resilience, but also a degree of plasticity. A supple network will absorb change rather than resist it.

To address these additional features of suppleness, we turn to a set of related measures, which provide a more fine-grained quantification of network structure. The first measure is called Fragmentation, and is defined as 1 minus the Krackhardt (1994) formula for network connectedness (Borgatti, 2006). Intuitively, the measure can be interpreted as the proportion of pairs of nodes that are unreachable from each other. In equation (4) below, $r_{ij} = 1$ if node i can reach node j by a path of any length and 0 otherwise. If all nodes are reachable from all others, then $F = 0$. If a graph contains only isolates, then $F = 1$.

$$F = 1 - \frac{2\sum_{i>j} r_{ij}}{n(n-1)} \tag{4}$$

The second measure is called wholeness (also known as the Component Ratio), and is defined as the number of components, s, minus one, divided by the number of nodes, n, minus one (Borgatti, 2006). $CR = 1$ when all nodes are isolates and $CR = 0$ when all nodes are part of the same component.

$$CR = \frac{s_c - 1}{n - 1} \tag{5}$$

As you can see from Figure 5, APL and size of the largest component are likely underestimating the true difference in structural change introduced by exogenous attack. The graphs generated with a preference for diversity maintain their global structure even as the attack continues concurrent with network growth. Moreover, the baseline graphs experience high volatility after change whereas the ER+D and BA+D graphs incorporate change while maintaining their original character. This contrast reveals suppleness in the (re)generating process when a preference for diversity is present.

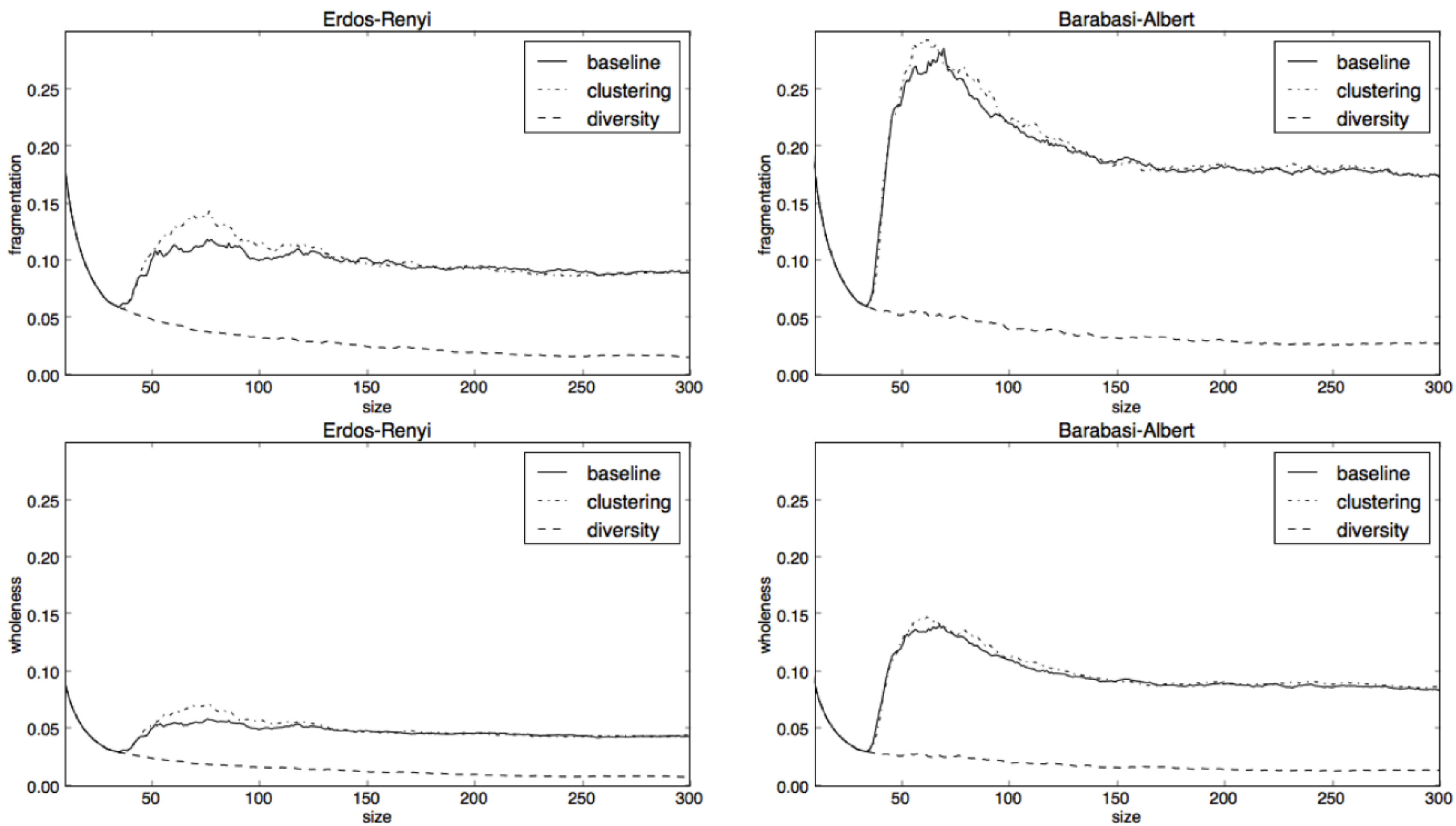

Fig. 5. Fragmentation and Wholeness with attacks starting at n=30. All values are averaged across 100 random trials.

8 Discussion

Networks serve as a platform for nearly all organized social and economic activity. The continuity and performance of many social and organizational phenomena relies, in turn, on the stability of network structures in the face of external pressures. In the extant literature on complex networks, these pressures take two forms, decay and attack, while stability is captured via measures of resistance and resilience.

Decay is the occurrence of small perturbations to a network, introduced by the rewiring of a small number of ties typically among more peripheral nodes, but otherwise chosen at random. Attack is the targeted removal of a central node, or a small number of more central nodes, based on their position or other attributes. Decay and attack each have many analogs in the evolution of real industrial networks. Instantiations of decay at the organizational level include the dissolution of ties concomitant to the end of a joint project, or the reshuffling of ties that often follows the movement of individual employees–these individuals may take their clients, grants, or other projects, from one firm to another. Attack is exemplified when larger, more central firms become targets for takeover by outside entities, but can also occur if critical lines of resources are cut off exogenously.

Resistant structures remain rigid in their global properties in spite of exogenous changes. Resilient networks are elastic, at first deforming to accommodate exogenous change, but quickly rebounding so that global properties are restored to prior levels.Such definitions of stability are apt enough when the imposed changes are one-offs–idiosyncratic, independent, and transient. Yet, in many social and organizational settings,networks must deal with changes that are more integral, interdependent, and long lasting. Environments undergo fundamental transformation (Haveman, 1992; Haveman *et al.*, 2001), technological equilibriums punctuate (Tushman & Anderson, 1986; Anderson & Tushman, 1990),suitors persist (Loch & Huberman, 1999) and resources partition (Boone *et al.*, 2002; Carroll & Swaminathan, 2000). Resistance and resilience may leave the network intact, but obsolete, unfit,or otherwise frail.

We introduced the concept of suppleness to capture the ability of a network to incorporate change without losing its structural integrity and defining characteristics.Some networks are brittle–they do not suffer change gracefully, such as when subjected to decay or attack (Callaway *et al.*, 2000; Holme *et al.*, 2002). When key nodes or edges are removed, brittle network structures may degrade rapidly, break apart, or even collapse. Small-worlds become big places (Watts & Strogatz, 1998; Holme *et al.*, 2002), structures of accumulated advantage collapse (Holme, 2002), and connected cavemen turn their backs on all encampments but their own (Holme & Kim, 2002). For example, Onnela *et al.* (2007) describe a mobile communications database in which social networks quickly fragment when ties are removed. The resultant structure traps information in silos, thus preventing the flow of useful resources between clusters. Uzzi (1997) provides and even more intimate account using the context of interorganizational ties. Cohesion improves performance up to a threshold after which the cluster of firms becomes vulnerable to exogenous shock.

Other networks are supple, responding less dramatically to change. Supple networks absorb shocks, such as when key nodes or edges are removed, while maintaining their structural character and integrity. Supple networks have "give." That is, the numerical properties of supple networks can change, within some limits, while the network maintains its function–whether it be providing novel ideas, social support, or concealment (Fleming *et al.*, 2007; Tiwana, 2008).

Many real-world networks seem to exhibit suppleness. For example, Powell *et al.* (2005) studied the network of overlapping and redundant strategic alliances in the biotech industry, which remained largely unaffected by numerous consolidating events (e.g. mergers, acquisition and exit). It may well be that social systems of a certain complexity require diversity as a precondition for their continued existence. However, such an argument is not new. Comte (1852) suggests that it is "the continuous repartition of different human endeavors which especially constitutes social solidarity..."–an idea leveraged and refined by Durkheim (1893/1997). Simply stated, society is more flexible, more plastic, indeed more supple, as the diversity produced by a division of labor is allowed to flourish.

Yet, archetypes of supple graphs are lacking. Individual-level rules of attachment result in structures that are characterized both by the topology of global or socio-centric worlds and the functioning of local or ego-centric neighborhoods. The micro and macro, or local and global, are considerably indeterminate of one another. However, certain archetypes have garnered considerable attention precisely for the interplay between local and global that they reveal. The algorithms of Watts and Strogatz (1998), Holme and Kim (2002), and Barabási and Albert (1999) serve as constructivist theories of social structure, and yet each of the resultant networks is often more brittle than the social system it professes to describe. Granted, ours is a theory of nascent network structure. Social and interorganizational networks may be subject to broader logics of attachment as certain processes of legitimacy and industry lifecycle play out (Klepper, 1996; Hannan & Freeman, 1977).

In order to achieve the the small-world structure developed by Watts and Strogatz (1998), it suffices to maintain just one cross-cutting (weak) tie to distant network clusters. This formulation is brittle in the most obvious sense–it lacks redundancy. Moreover, there is no motivation for creating additional cross-cutting ties implied by the theory. The popularity bias described by Barabási and Albert (1999) suffers a similar limitation. During the formative stages of the network, there is no motivation to spread ties amongst several founding nodes. Instead, just a few "lucky" individuals receive the majority of incoming connections. Thus, the most popular nodes also serve as easy targets for exogenous attack.

Recent empirical findings seem to suggest that other processes may be at play (Hargadon & Sutton, 1997; Powell *et al.*, 2005; Shemtov, 2003). Indeed, actors distinguish between resources. Consider a biotech firm's partner network. If that firm currently has only one partner–a university for instance–then they are apt to pursue subsequent ties with partners of a different form. For instance, they might pursue ties with a venture capitalist, pharmaceutical or law firm–each of which brings a complementary set of resources–rather than additional ties to another university regardless of how prominent that university is. This argument can be extended to the tie-seeking behavior of both individuals and firms. Clearly, it is optimal to explore a variety of options early–a point at which uncertainty is highest–as a means to develop preferences used in later encounters (March, 1991). For instance, Koput and Gutek (2010) show that women who establish cross-gender ties early in their careers are more successful than those who attempted to establish such ties later on. Given these tendencies, the early stages of any social network algorithm should consider a bias in favor of categorical diversity.

This emphasis on nodes with few ties is addressed by our model directly. During the formative stages of a network, the influence of popularity and age is attenuated by the impulse of new nodes to build a diverse portfolio of ties. This renewal process ensures less variance in degree overall and more redundant cross-cutting ties. Through this process, we argued, and demonstrated, that a preference for diversity in tie formation can generate supple networks–and can add suppleness to networks when introduced into other models of network formation.

References

Albert, Réka, Jeong, Hawoong, & Barabási, Albert-László. (2000). Error and attack tolerance of complex networks. *Nature*, **406**(6794), 378–382.

Anderson, Philip, & Tushman, Michael L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative science quarterly*, 604–633.

Baker, Wayne E., & Faulkner, Robert R. (1993). The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American sociological review*, 837–860.

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

Bishop, Matt, Carvalho, Marco, Ford, Richard, & Mayron, Liam M. (2011). Resilience is more than availability. *Pages 95–104 of: Proceedings of the 2011 workshop on new security paradigms workshop*. ACM.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. Free Press New York.

Boland, Wendy Attaya, Brucks, Merrie, & Nielsen, Jesper H. (2012). The attribute carryover effect: What the "runner-up" option tells us about consumer choice processes. *Journal of consumer research*, **38**(5), 872–885.

Boone, Christophe, Witteloostuijn, Arjen Van, & Carroll, Glenn R. (2002). Resource distributions and market partitioning: Dutch daily newspapers, 1968 to 1994. *American sociological review*, 408–431.

Borgatti, Stephen P. (2006). Identifying sets of key players in a social network. *Computational & mathematical organization theory*, **12**(1), 21–34.

Callaway, Duncan S., Newman, Mark EJ, Strogatz, Steven H., & Watts, Duncan J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical review letters*, **85**(25), 5468–5471.

Carroll, Glenn R., & Swaminathan, Anand. (2000). Why the microbrewery movement? organizational dynamics of resource partitioning in the us brewing industry1. *American journal of sociology*, **106**(3), 715–762.

Clauset, Aaron, Shalizi, Cosma Rohilla, & Newman, Mark EJ. (2009). Power-law distributions in empirical data. *Siam review*, **51**(4), 661–703.

Comte, Auguste. (1852). *Cour de philosophie positive*. Vol. 1. Borrani et Droz.

Durkheim, Emile. (1893/1997). *The division of labor in society*. Simon and Schuster (Original work published 1893).

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Magyar tud.akad.mat.kutató int.közl*, **5**, 17–61.

Fleming, Lee, Mingo, Santiago, & Chen, David. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative science quarterly*, **52**(3), 443–475.

Gibbs, Jack P., & Martin, Walter T. (1962). Urbanization, technology, and the division of labor: International patterns. *American sociological review*, 667–677.

Girvan, Michelle, & Newman, Mark EJ. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, **99**(12), 7821–7826.

Granovetter, Mark S. (1973). The strength of weak ties. *American journal of sociology*, 1360–1380.

Hannan, Michael T., & Freeman, John. (1977). The population ecology of organizations. *American journal of sociology*, 929–964.

Hargadon, Andrew, & Sutton, Robert I. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, 716–749.

Haveman, Heather A. (1992). Between a rock and a hard place: Organizational change and performance under conditions of fundamental environmental transformation. *Administrative science quarterly*, 48–75.

Haveman, Heather A., Russo, Michael V., & Meyer, Alan D. (2001). Organizational environments in flux: The impact of regulatory punctuations on organizational domains, ceo succession, and performance. *Organization science*, **12**(3), 253–273.

Herfindahl, Orris Clemens. (1950). Concentration in the steel industry. *Concentration in the steel industry*.

Holme, Petter. (2002). Edge overload breakdown in evolving networks. *Physical review e*, **66**(3), 036119.

Holme, Petter, & Kim, Beom Jun. (2002). Growing scale-free networks with tunable clustering. *Physical review e*, **65**(2), 026107.

Holme, Petter, Kim, Beom Jun, Yoon, Chang No, & Han, Seung Kee. (2002). Attack vulnerability of complex networks. *Physical review e*, **65**(5), 056109.

Klepper, Steven. (1996). Entry, exit, growth, and innovation over the product life cycle. *The american economic review*, 562–583.

Koput, Kenneth William, & Gutek, Barbara A. (2010). *Gender stratification in the it industry: Sex, status and social capital*. Edward Elgar Publishing.

Krackhardt, David. (1994). Graph theoretical dimensions of informal organizations. *Computational organization theory*, **89**(112), 123–140.

Loch, Christoph H., & Huberman, Bernardo A. (1999). A punctuated-equilibrium model of technology diffusion. *Management science*, **45**(2), 160–177.

March, James G. (1991). Exploration and exploitation in organizational learning. *Organization science*, **2**(1, Special Issue: Organizational Learning: Papers in Honor of (and by) James G. March), pp. 71–87.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415–444.

Onnela, J-P, Saramäki, Jari, Hyvönen, Jorkki, Szabó, György, Lazer, David, Kaski, Kimmo, Kertész, János, & Barabási, A-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, **104**(18), 7332–7336.

Owen-Smith, Jason, Buhr, Helena, & Funk, Russel J. (2014). Network portfolios and the authenticity premium: Market value across five high-technology sectors. *Working paper*.

Pachucki, Mark A., & Breiger, Ronald L. (2010). Cultural holes: Beyond relationality in social networks and culture. *Annual review of sociology*, **36**, 205–224.

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.

Powell, W. W., Koput, K. W., Smith-Doerr, L., & Owen-Smith, J. (1999). Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. *Research in the sociology of organizations*, **16**(1), 129–159.

Powell, W. W., White, D. R., Koput, K. W., & OwenSmith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences1. *American journal of sociology*, **110**(4), 1132–1205.

Shemtov, Ronit. (2003). Social networks and sustained activism in local nimby campaigns. *Chap. 2, pages 215–244 of: Sociological forum*, vol. 18. Springer.

Shore, Jesse, Chu, Catherine J., & Bianchi, Matt T. (2013). Power laws and fragility in flow networks. *Social networks*, **35**(1), 116–123.

Stinchcombe, A. L. (1965). Organizations and social structure. *Handbook of organizations*, **44**(2), 142–193.

Taylor, Verta. (1989). Social movement continuity: The women's movement in abeyance. *American sociological review*, 761–775.

Tellis, Gerard J. (1988). Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of marketing research*, **25**(2), 134–144.

Tiwana, Amrit. (2008). Do bridging ties complement strong ties? an empirical examination of alliance ambidexterity. *Strategic management journal*, **29**(3), 251–272.

Tushman, Michael L., & Anderson, Philip. (1986). Technological discontinuities and organizational environments. *Administrative science quarterly*, 439–465.

Uzzi, Brian. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, 35–67.

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

Wimmer, Andreas, & Lewis, Kevin. (2010). Beyond and below racial homophily: Erg models of a friendship network documented on facebook1. *American journal of sociology*, **116**(2), 583–642.