AFRL-AFOSR-VA-TR-2015-0394

Network Dynamics: Modeling And Generation Of Social Networks

Sidney Redner
TRUSTEES OF BOSTON UNIVERSITY
1 SILBER WAY
BOSTON, MA 02215-1390

11/23/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
23-11-2015

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
01-08-2012 to 31-07-2016

4. TITLE AND SUBTITLE
Network Dynamics: Modeling And Generation Of Very Large Heterogeneous Social Networks

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0391

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Sidney Redner

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TRUSTEES OF BOSTON UNIVERSITY
1 SILBER WAY
BOSTON, MA 02215-1390 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
One major achievement was the construction of redirection algorithms to efficiently generate large networks with prescribed degree characteristics.
A hindered redirection algorithm was shown to reproduce sublinear preferential attachment. Conversely, enhanced redirection leads to
highly-dispersed networks that contain multiple macrohubs (degree a finite fraction of the number of network nodes) and exhibit non-extensive
scaling.
The average number of distinct degrees that appear in a finite network was found to grow algebraically with network size and the underlying
distribution is a universal Gaussian. A choice-driven network growth mechanism was formulated in which a new node first identifies a set of target
nodes and attaches to either the target with the largest degree (greedy choice), or the target whose degree is not the largest (meek choice). The
resulting network exhibits a non-universal power-law degree distribution.

15. SUBJECT TERMS
network, social, graphs

16. SECURITY CLASSIFICATION OF:

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UU		Sidney Redner
					19b. TELEPHONE NUMBER *(Include area code)*
					617-353-2618

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release.

NETWORK DYNAMICS: MODELING AND GENERATION OF VERY LARGE HETEROGENEOUS SOCIAL NETWORKS

Final Report, November 12, 2015

Sidney Redner (redner@santafe.edu), Santa Fe Institute, Santa Fe, NM 87501
Paul Krapivsky (paulk@bu.edu), Dept. of Physics, Boston Univ., Boston, MA 02215

Publications with DARPA Support

1. "Emergence of Clustering in an Acquaintance Model without Homophily", U. Bhat, P. L. Krapivsky, and S. Redner, J. Stat. Mech. P11035 (2014).
2. "Highly Dispersed Networks Generated by Enhanced Redirection", A. Gabel, P. L. Krapivsky, and S. Redner, J. Stat. Mech. P04009 (2014).
3. "Choice-Driven Phase Transition in Complex Networks" P. L. Krapivsky and S. Redner, J. Stat. Mech. P04021 (2014).
4. "Highly Dispersed Networks by Enhanced Redirection, A. Gabel, P. L. Krapivsky, and S. Redner, Phys. Rev. E **88**, 050802 (2013).
5. "Distinct Degrees and Their Distribution in Complex Networks", P. L. Krapivsky and S. Redner, J. Stat. Mech. P06002 (2013).
6. "Sublinear but Never Superlinear Preferential Attachment by Local Network Growth", A. Gabel and S. Redner, J. Stat. Mech. P02043 (2013).

Summary

We constructed redirection algorithms to efficiently generate networks with prescribed characteristics. In redirection, each newly introduced node either links to a random target (probability $1-r$) or to the parent of the target (probability r). When r is fixed, this gives linear preferential attachment—the attachment rate to a node is a linear function of its degree. When r is a decaying function of the parent degree, redirection reproduces sublinear preferential attachment. When r an increasing function of the parent degree, highly-dispersed networks arise. These contain multiple macrohubs—whose degree is a finite fraction of the number of network nodes and exhibit non-extensive scaling of the degree distribution.

We found intriguing properties in the number of distinct degrees in a typical realization of a complex network. The average number of distinct degrees grows algebraically with network size, with an exponent that is the reciprocal of the degree distribution exponent. The distribution of distinct degrees is a universal Gaussian. We also investigated choice-driven network growth where a new node first provisionally selects set of target nodes and attaches to either: (i) the target with the largest degree (greedy choice), and (ii) a target whose degree is *not* the largest (meek choice). The resulting network may have either:

(i) a non-universal power-law degree distribution or (ii) a single macroscopic hub. At the transition between cases (i) and (ii) the degree distribution decays as $(k \ln k)^{-2}$

We generated heterogeneous social networks from homogeneous interaction rules. Heterogeneity emerges dynamically as the network grows, rather than being imposed explicitly by homophily or by heterogeneous individual agents. The networks that result from our algorithm resemble real social networks, such as those available from Facebook. Finally, we constructed *dense* networks through a simple node copying mechanism. These networks exhibit many unusual features, including multiple structural phase transitions in the densities of cliques (small complete graphs), and a log-normal degree distribution in the dense regime.

Preliminary: Redirection Algorithm

In redirection, the addition of each new node **n** to the network occurs by (Fig. 1(a)):

1. Node **n** randomly selects an arbitrary "target" node **x** from the network.
2. With probability $1 - r$, with $0 < r < 1$, node **n** attaches to **x**.
3. With probability r, node **n** links to the *parent node* **y** of **x**.

This rule generates a linear preferential attachment network, in which the induced rate A_k at which a new node attaches to an existing node of degree k is given by

$A_k = k + \lambda$, with $\lambda = \frac{1}{r} - 2$, namely, this algorithm reproduces shifted linear preferential attachment [1]. Thus a *purely local rule* generates a network with a *globally-defined* design criterion. This algorithm is extremely efficient, as the time required to build a network of N nodes grows linearly with N. For $r = 0$, the classic random recursive tree (RRT) arises, in which the fraction of nodes of degree k, $n_k = N_k/N$, asymptotically scales as $n_k \simeq 2^{-k}$. For $r > 0$, a non-universal power-law degree distribution $n_k \sim k^{-(3+\lambda)}$ arises. We emphasize that the degree distribution is sensitive to the additive factor λ in the attachment rate $A_k = k+\lambda$. Such a sensitivity contradicts conventional universality in critical phenomena, where such microscopic model details are irrelevant.

Hindered Redirection and Sublinear Preferential Attachment

In hindered redirection, we constructed networks in which the attachment rate A_k to a node of degree k is proportional to k^γ, with $\gamma < 1$. We achieved this growth law by augmenting redirection with local degree information [2]. In the original redirection algorithm, the probability of redirecting from an initial target to its parent is fixed. We extended this rule by specifying the redirection probability $r = r(a, b)$ to be a function of the degrees of the target and parent nodes, a and b, respectively (Fig. 1(b)). Let us define f_k as the total probability that an incoming link is redirected *from* a randomly-selected target node of degree k to the parent of the target. Similarly, we define t_k as the total probability that an incoming link is redirected *to* a parent node of degree k after the incoming node initially selected one of the child nodes of this parent. Formally, these probabilities are defined in terms of the

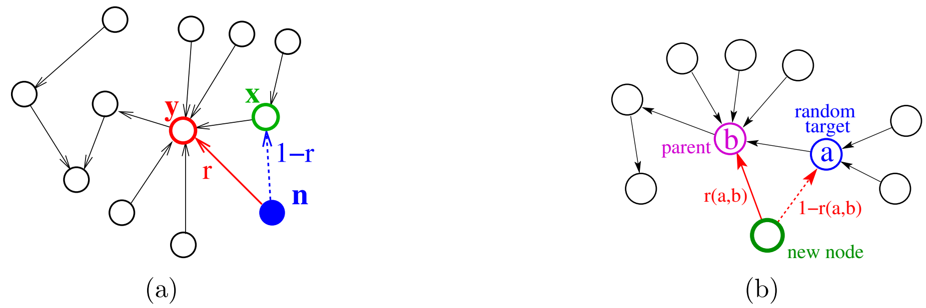


Figure 1: (a) Illustration of redirection. A new node **n** selects a target node **x** at random. With probability $1-r$, **n** links to this target (dashed arrow). With probability r, **n** links to **y**, the parent of **x** (thick solid arrow). (b) In hindered and in enhanced redirection, the redirection probability r is a function of the degrees of the initial target node and the parent node, a and b, respectively.

redirection probabilities by

$$f_k = \sum_{b\geq 1} \frac{r(k,b)N(k,b)}{N_k}\,, \qquad\qquad t_k = \sum_{a\geq 1} \frac{r(a,k)N(a,k)}{(k-1)N_k}\,, \tag{1}$$

where $N_k = \sum_{b\geq 1} N(k,b)$ and $N(a,b)$ is the correlation function that specifies the number of nodes of degree a that have a parent of degree b. Thus f_k is the redirection probability averaged over all N_k possible target nodes of degree k. Likewise, since each node of degree k has $k-1$ children, there are $(k-1)N_k$ possible target nodes whose redirection probabilities are averaged to give t_k.

In terms of these probabilities f_k and t_k, the master equation that governs the evolution of the degree distribution N_k is

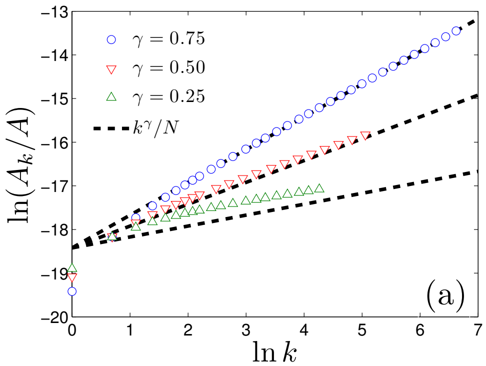
$$\frac{dN_k}{dN} = \frac{(1-f_{k-1})N_{k-1} - (1-f_k)N_k}{N} + \frac{(k-2)t_{k-1}N_{k-1} - (k-1)t_k N_k}{N} + \delta_{k,1}. \tag{2}$$

The first ratio corresponds to realizations of the growth process for which the incoming node actually attaches to the initial target. For example, the term $(1-f_k)N_k/N$ gives the probability that one of the N_k target nodes of degree k is randomly selected and that the link from the new node is *not* redirected away from this target. Similarly, the second ratio corresponds to instances in which the link to the target node *is* redirected to the parent. For example, the term $(k-1)t_k N_k/N$ gives the probability that one of the $(k-1)N_k$ children of nodes of degree k is chosen as the target and that the new node *is* redirected. The term $\delta_{k,1}$ accounts for the newly-added node of degree 1.

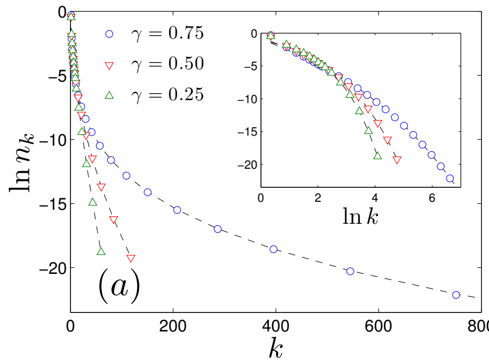
Rearranging terms, we express the master equation (2) in the generic form,

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k,1}\,, \tag{3}$$

where N_k is the number of nodes of degree k and A_k is the rate at which a new node attaches to a target node of degree k. with attachment rate given by $A_k/A = [(k-1)t_k + 1 - f_k]/N$.



(a)



(b)

Figure 2: (a) Simulated attachment probabilities A_k/A versus k from generalized redirection. (b) Degree distribution n_k versus k for generalized redirection with $r(a,b) = b^{\gamma-1}$ for $\gamma = 0.75$ ($\circ$), $\gamma = 0.5$ ($\bigtriangledown$), and $\gamma = 0.25$ ($\bigtriangleup$). Inset: the same data on a double logarithmic scale.

When the redirection probability is constant, we have $f_k = t_k = r$, and the expected linear dependence $A_k \sim k$ is recovered. Note that the large-k behavior of the attachment rate is $A_k \sim k\, t_k$. Thus a redirection probability $r(a,b)$ for which t_k is a decreasing function of k will asymptotically correspond to sublinear preferential attachment. We focused on the specific example where $r(a,b) = b^{\gamma-1}$, with $0 < \gamma < 1$. Because r depends only on the degree of the parent node, Eq. (1) reduces to $t_k = k^{\gamma-1}$. Using this form of t_k in the attachment rate asymptotically gives sublinear preferential attachment, with $A_k \sim k^\gamma$.

To test the prediction that $A_k \sim k^\gamma$ with $\gamma < 1$, we simulated networks that are grown to $N = 10^8$ nodes by generalized redirection. Once a network reaches this size, we measured the probability that attachment to a node of degree k actually occurs by systematically making 'test' attachments to each node of the network according to generalized redirection. The term test attachment means that the network is returned to its original state after each such event. We count all test events that ultimately lead to attachment to a node of degree k. Dividing this number of events by the total number of nodes N gives the attachment probability to nodes of degree k, $A_k N_k/A$.

Our simulations show that the A_k grows sublinearly with k (Fig. 2(a)). The agreement is best for γ less than, but close to 1, where the degree distribution is sufficiently broad that meaningful statistical tests can be performed, while for small γ, the asymptotic behavior is contaminated by progressively more slowly-decaying sub-asymptotic corrections terms in the expression for the attachment rate. Our simulated degree distribution is also close to the analytic result [1]

$$n_k = \frac{\mu}{A_k} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j}\right)^{-1} \sim k^{-\gamma} \exp\left[-\frac{\mu}{1-\gamma}\; k^{1-\gamma}\right]. \tag{4}$$

Our primary result is that sublinear preferential attachment networks can be generated with a similar algorithmic simplicity as in shifted linear preferential attachment.

Enhanced Redirection and Highly-Dispersed Networks

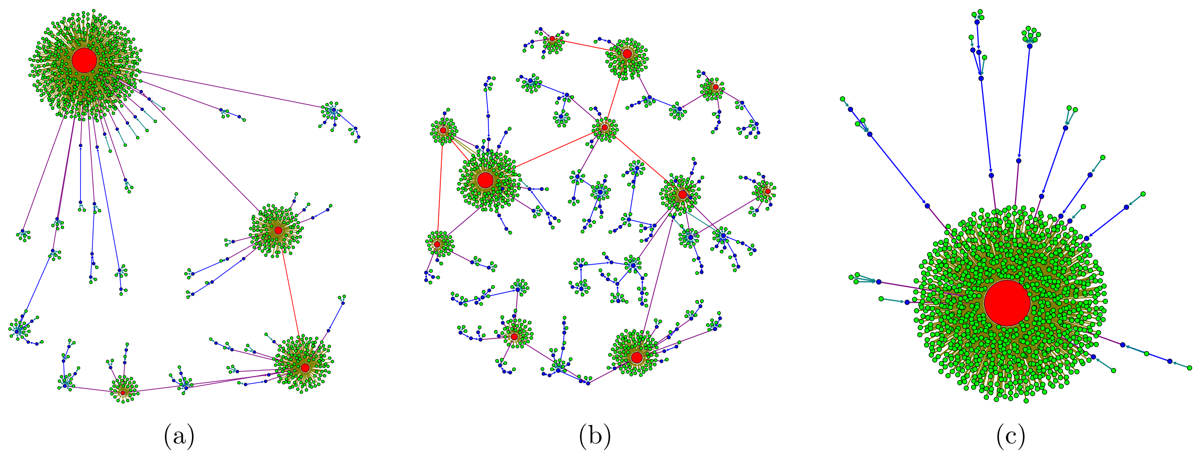


Figure 3: Example enhanced redirection networks of 10^3 nodes for $\lambda = \frac{3}{4}$. (a) Maximum degree $k_{\max} = 548$, $\mathcal{C} = 66$ core (degree ≥ 2) nodes, and maximum depth (root has depth 0, its children have depth 1, etc.) $D_{\max} = 10$. (b) $k_{\max} = \mathcal{C} = 154$, $D_{\max} = 12$ (the smallest $k_{\max}$ out of 10^3 realizations). (c) $k_{\max} = 963$, with $\mathcal{C} = 23$ and $D_{\max} = 6$ (the largest $k_{\max}$ out of 10^3 realizations). White: nodes of degree 1; black: degrees 2–20; gray (red): degree > 20.

In *enhanced redirection*, the redirection probability r is an increasing function of the parent degree b with $r \to 1$ as $b \to \infty$ (Fig. 1). Here, for simplicity, we focus on the situation where each node has out-degree equal to 1, and thus a unique parent [3, 4]; our approach extends to networks where each node has out-degree greater than 1, so that closed loops arise [4]. For convenience, we choose the initial condition of a single root node of degree 2 that links to itself. The root is thus both its own parent and its own child. To grow the network, nodes are introduced one by one and each new node $\mathbf{n}$ first picks a random target node (of degree a), and then (Fig. 1(b)):

(i) either $\mathbf{n}$ node attaches to the target with probability $1 - r(a, b)$;

(ii) or $\mathbf{n}$ attaches to the parent (with degree b) of the target with probability $r(a, b)$.

For enhanced redirection, we focused on the redirection probability $r(a, b) = 1 - b^{-\lambda}$ with $\lambda > 0$, but other forms for which $r(a, b) \to 1$ as $b \to \infty$ give similar results. This redirection rule gives rise to networks with the following intriguing properties (Fig. 3):

1. ***Multiple Macrohubs*:** Macrohubs are nodes whose degrees are a finite fraction of N. While macrohubs arise in other models [1, 5–7], the resulting networks are singular.

For superlinear preferential attachment, where $A_k \sim k^\gamma$ with $\gamma > 1$ [1, 5] and in the fitness model, where the attachment rate is proportional to both the degree k and fitness of the target [7, 8], a single macrohub arises that is connected to almost all other nodes of the network. In contrast, enhanced redirection networks are highly disperse (Fig. 3), with interconnected hub-and-spoke structures that are reminiscent of airline route networks [6, 9–12].

2. ***Non-extensivity*:** In sparse networks, the degree distribution is extensive, with the number of nodes of degree k, N_k, proportional to N. In linear preferential attachment, for example, the degree distribution has an algebraic tail, $N_k \sim N/k^\nu$ for $k \gg 1$, with $\nu > 2$. In contrast, enhanced redirection leads to the non-extensive scaling

$$N_k \sim \frac{N^{\nu-1}}{k^\nu} \qquad \text{with} \quad \nu < 2. \tag{5}$$

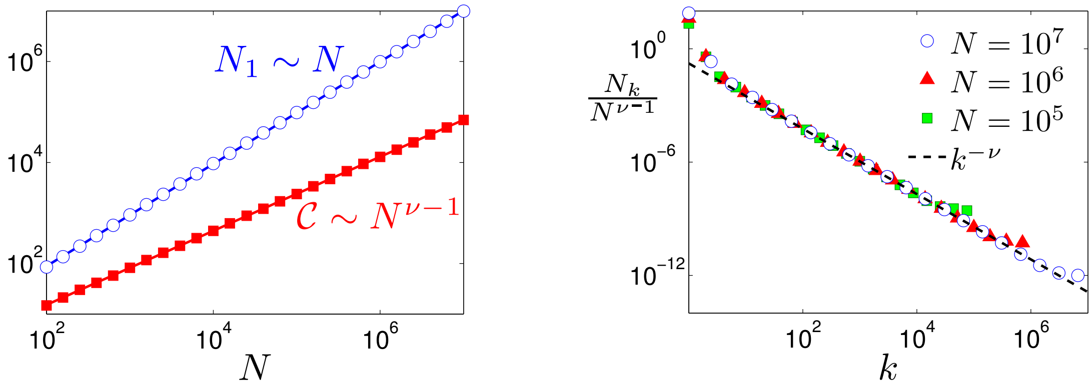
   The allowed range of the exponent ν is key. While past empirical studies have observed networks with degree exponent in the range $1 < \nu < 2$ [13], the range $1 < \nu < 2$ is mathematically inconsistent for sparse networks because it leads to a divergent average degree as $N \to \infty$ if the degree distribution obeys the standard scaling assumption $N_k \sim Nk^{-\nu}$. This dilemma is resolved if the degree distribution is not extensive. In enhanced redirection, almost all nodes have degree 1 (leaves), while, the number $\mathcal{C} \equiv N - N_1$ of core nodes (nodes with degree > 1) grows *sub-linearly* with N; that is, as $N^{\nu-1}$ with exponent ν in the range $1 < \nu < 2$. All N_k with $k \geq 2$ also grow as $N^{\nu-1}$. This anomalous scaling can be summarized as:

$$\mathcal{C} = N - N_1 \simeq c_1 N^{\nu-1}\,, \qquad N_k \simeq c_k N^{\nu-1} \quad \text{for} \quad k \geq 2 \tag{6}$$

   where c_k are constants. This scaling leads to a finite average degree $\langle k \rangle$ without imposing an artificial cutoff in the degree distribution.

3. ***Lack of Self-Averaging*:** Different network realizations in enhanced redirection are visually diverse (Fig. 3). The number of nodes of fixed degree, N_k with any $k \geq 2$, and the number of core nodes $\mathcal{C}$, vary significantly between realizations. For instance, the ratio of the deviation to the average, $\sqrt{\langle \mathcal{C}^2 \rangle - \langle \mathcal{C} \rangle^2}/\langle \mathcal{C} \rangle$, converges to a positive constant as $N \to \infty$, so that the number of core nodes do not self-average. In contrast, preferential attachment networks do self-average, as the relative deviations in N_k or $\mathcal{C}$ systematically decrease as N increases [14].

Our simulations show that the degree distribution (Fig. 4) scales anomalously, as given by Eq. (6). The exponent ν depends on the redirection parameter λ, but is always less than 2 (Fig. 4) so that the degree distribution decays very slowly in k. Because $\nu < 2$, Eq. (6) implies that the number of nodes of degree 1 grow more rapidly with N than core nodes. Thus, visually, a typical network in enhanced redirection is dominated by its leaves.



(a) (b)

Figure 4: (a) N_k versus N and (b) $N_k/N^{\nu-1}$ versus k for enhanced redirection with $\lambda = \frac{3}{4}$ and $\nu = 1.73$ (determined numerically). Data are based on 10^4 realizations, with equally-spaced bins on a logarithmic scale in (b). The lines in (a) show the prediction of Eq. (6), while the line in (b) shows the k dependence from the numerical solution of the master equations.

Distinct Degrees

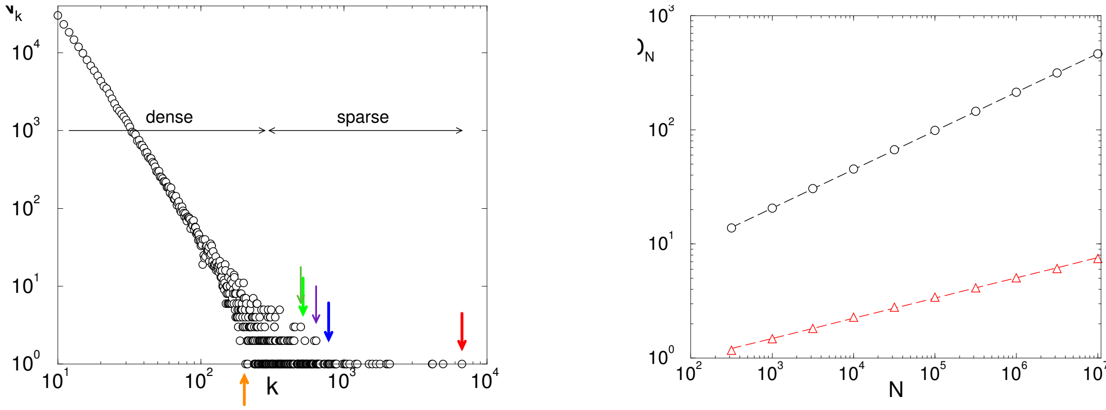
An unexpected byproduct of our research on generative models is that we found that the number of distinct degrees that exist in a typical realization of a complex network display intriguing statistical properties [15]. Our basic result: for a single realization of a preferential attachment network of N nodes that is drawn from an ensemble in which $N_k \sim N/k^\nu$ for $k \gg 1$, the number of distinct degrees grows as $N^{1/\nu}$.

To determine the number of distinct degrees that appear in a typical realization of a large network, first notice that for k in the range $k \leq K = (NR)^{1/\nu}$, there are many nodes with such degrees, so that $N_k \geq 1$. In this dense regime of the degree distribution (Fig. 5), all degrees with $k < K$ are present. This range therefore gives a contribution of $(NR)^{1/\nu}$ to D_N. In the complementary sparse range of $k > K$, we estimate the number of distinct degrees, by integrating the degree distribution for $k > K$. Adding the contributions from the dense and sparse regimes gives

$$D_N^{\rm naive} = \frac{\nu}{\nu - 1}\, K, \qquad\qquad K = (NR)^{1/\nu}\,. \tag{7}$$

While the N-dependence is correct, $D_N \sim N^{1/\nu}$, the amplitude is wrong. We obtained a better estimate by assuming that the probability distribution for the number of nodes of each degree k is a Poisson distribution with average value N_k given by $N_k \sim N/k^\nu$. Then $P_k \equiv \text{Prob}[(\#\text{ nodes of degree } k) \geq 1] = 1 - \exp(-N_k)$. Using this property, we found the more accurate estimate $D_N = \sum_{k\geq 1}\left[1 - e^{-N_k}\right]$ which leads, in the large N limit, to

$$D_N \simeq \Gamma\big(1 - \tfrac{1}{\nu}\big)\, (RN)^{1/\nu} \tag{8}$$



(a) (b)

Figure 5: (a) Number of nodes of degree $k \geq 10$ for a single linear preferential attachment network of $N = 10^7$ nodes. The largest degree is $k_{\max} = 6693$, $D_N = 465$, the last doublet occurs at $k = 782$, the last dimer at $k = 641$, the last triplet at $k = 518$, the last trimer at $k = 500$, and the first hole at $k = 201$ (arrows). (b) The average number of distinct degrees D_N versus N. The upper curve ($\circ$) corresponds to the shifted linear preferential attachment rate $A_k = k - \frac{1}{2}$. Here $D_N = BN^{2/5}$, with $B = (3/2)^{2/5}\pi^{-1/5}\Gamma(3/5) = 1.393019\ldots$. The lower curve ($\triangle$) corresponds to $A_k = k$. Each data point represents an average over 10^4 realizations. The dashed lines correspond to our theoretical predictions.

for networks whose degree distribution has the algebraic tail $N_k \simeq NRk^{-\nu}$ for $k \gg 1$, and R is a constant of the order of 1.

For strictly linear preferential attachment, $R = 4$ and $\nu = 3$, so that $D_N = BN^{1/3}$, with $B = 2^{2/3}\Gamma\left(\frac{2}{3}\right) = 2.149528\ldots$. For this simplest model, we also studied the probability distribution $\Pi(D_N)$ of distinct degrees and found that it has the Gaussian form, with standard deviation growing as $\sqrt{\langle D_N^2\rangle - \langle D_N\rangle^2} \sim N^{1/6}$.

Statistical characteristics of the sparse tail of the degree distribution exhibit interesting universal behaviors. This includes the location of the first "hole" in the degree distribution—the smallest degree value for which $N_k = 0$. Using probabilistic arguments, we found that the first hole is asymptotically located at the degree value

$$\langle h_1 \rangle \simeq \left(\frac{\nu N R}{\ln N}\right)^{1/\nu} . \qquad (9)$$

We also determined the location of the last doublet, the largest two consecutive k values for which $N_k > 0$, and the last dimer, the largest k value for which $N_k = 2$ (Fig. 5). Starting with degree 1, the degree distribution first consists of a long string of consecutive "occupied" degrees until the first hole, followed by a second string until the second hole, etc. As the degree increases, these strings become progressively shorter and above a certain threshold all remaining strings are singlets. For a large network, the last string that is not a singlet will almost certainly be a doublet (with probability approaching 1 as $N \to \infty$). Using probabilistic reasoning, the average position of this last doublet is given by

$$\langle \delta \rangle = C(RN)^{1/(\nu-1/2)}\,. \tag{10}$$

We also quantifies other characteristics of the sparse tail of the degree distribution, such as the location of the last higher multiplets, and the last dimer, the last trimer, etc.

Our results apply to other heavy-tailed integer-valued distributions. The universal Gaussian nature of the distribution of distinct elements arising in a given realization suggests a new yet-to-be-proved central-limit theorem that underlies these results.

Choice-Driven Phase Transition

A remarkable outcome of network research is that realistic networks can be build via random algorithms. The lack of any *central* planning is a virtue in many situations, but the lack of local choice can be problematic [16]. We addressed this issue by the following choice-driven growth rule [17]: a set of target nodes is first selected, and a new node attaches to the optimal target. We defined optimality in terms of node degree. Specifically, the new node attaches to the target with the largest degree; this defines the "greedy" choice model (Fig. 6).

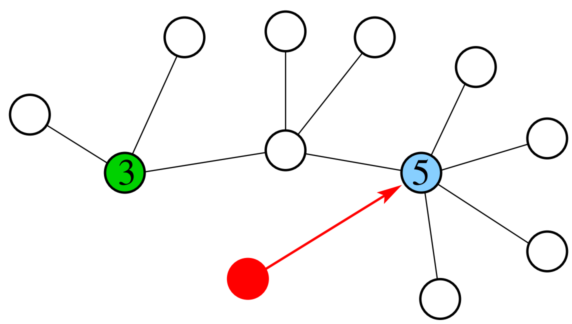


Figure 6: Illustration of network growth by greedy choice out of two alternatives. Two nodes in the network are selected according to preferential attachment. A new node attaches to the target with the larger degree, in this case, degree 5.

The simples non-trivial case is choice from two alternatives—two nodes are provisionally selected, and the new node links to the node with with maximal degree (with random choice if both selected nodes have equal degrees). The initial selection of the target is made according to shifted linear preferential attachment. As we already know, this is efficiently done through the redirection mechanism with fixed redirection probability r. We find that the degree distribution for greedy choice has the algebraic tail, $n_k \sim k^{-\nu_2}$ for $k \gg 1$, where the decay exponent is given by

$$\nu_2(r) = \begin{cases} 1 + 1/(2r) & 0 < r < \frac{1}{2}\,, \\ 1 + 1/(2-2r) & \frac{1}{2} < r < 1. \end{cases} \tag{11}$$

and the subscript refers to greedy choice from two alternatives. The two forms for $\nu_2(r)$ coincide when $r = \frac{1}{2}$, which corresponds to the strictly linear preferential attachment. In this case, the degree distribution acquires an additional logarithmic correction, $n_k \simeq 4(k \ln k)^{-2}$. The most interesting feature is the emergence of a macrohub when $\frac{1}{2} < r < 1$. The degree of the macrohub is hN with $h = \frac{2r-1}{r^2}$.

These results generalize to greedy choice from $p \geq 2$ of alternatives. Now a macrohub with degree $h(r)N$ emerges when the redirection probability exceeds $r_c = \frac{1}{p}$, where

$$h = 1 - (1 - hr)^p \tag{12}$$

The degree distribution has an algebraic tail with exponent

$$\nu_p(r) = \begin{cases} 1 + 1/(pr) & 0 < r < \frac{1}{p}\,, \\ 1 + 1/(pr[1 - rh]^{p-1}) & \frac{1}{p} < r < 1. \end{cases}$$

and $h = h(r)$ implicitly determined by (12).

We also investigated attaching to a node in the target set whose degree is *not* the largest. For a target set of p nodes, there are $p - 1$ possible such choices—to the 2^{nd}-largest degree node, the 3^{rd} largest, ..., to the smallest-degree node. These *meek* choice models all exhibit a double-exponential degree distribution of the form $\exp(-\text{const.} \times e^k)$. The maximal degree is of the order of $\log_m \log_m N$ in the specific case where a new node attaches to the m^{th}-largest degree out of a target set of p nodes. For the meek choice from two alternatives, for instance, in simulations of 50 realizations of networks grown to 10^8 nodes, the largest observed degree was only 9. Thus meek choice provides an powerful control strategy to keep the degree distribution homogeneous.

Spontaneous Clustering in a Homogeneous Acquaintance Model

We introduced an agent-based acquaintance model in which social links are created by processes with no explicit homophily [18]. In spite of this homogeneous interaction, highly-clustered social networks can arise. The crucial feature of our model is a variable triadic closure rate. Namely, when an agent introduces two unconnected friends, the rate at which these friends connect depends on the number of their mutual acquaintances. As this triadic rate is varied, the social network undergoes a dramatic clustering transition. Close to the transition, the network consists of a collection of well-defined communities. As a function of time, the network can also undergo an incomplete gelation transition, in which the gel, or giant cluster, does not constitute the entire network, even at infinite time. Some of the clustering properties of our model also arise, but more gradually, in Facebook networks. Finally, we constructed a more realistic variant in which network realizations can be constructed that quantitatively match Facebook networks.

In our acquaintance model there are two disparate ways that new social connections are made (see Fig. 7):

(a) *Direct connections.* An agent that possesses zero or one acquaintances directly links to a randomly-selected agent.

(b) *Triadic closure.* An agent with at least two acquaintances introduces a pair of them to each other. These two agents then make a link.

The distinction between zero/one and two or greater acquaintances is arbitrary and has been implemented for tractability. However, qualitatively similar phenomenology arises is the transition between direct and induced connections occurs that threshold value $k > 2$, rather than $k = 2$.

(a) (b)

Figure 7: Illustration of (a) direct and (b) induced linking. In (a), the selected node (solid circle) has degree zero and it directly links to another node (dashed line). In (b), a node (open circle) is first selected, and two acquaintances (shaded) of this selected node may link (dashed line).

These induced interactions are the crucial new feature of our model. By controlling their rate, highly clustered networks can arise. Moreover, as the network evolves, it can undergo an unusual gelation transition from a state where the social network consists of many small disjoint communities to a dense network in which most agents are interconnected. We formulated the two distinct mechanisms to determine the rates of direct and induced linking:

1. *Threshold-Controlled:* When two agents are selected as a result of induced linking, they connect if the ratio of the number of their mutual friends (inside the oval in Fig. 8) to the total number of friends of either agent equals or exceeds a specified threshold T.

2. *Rate-Controlled:* The direct and induced linking rates are 1 and R.

The rate at which triadic closure occurs can be either threshold controlled or explicitly rate controlled. Both versions lead to highly-clustered networks near a critical value of the relative rates of direct and triadic linking. In the parameter regime where the long-time network is highly clustered, there always remains a small population of marginal individuals that are part of isolated groups

Visualizations of small networks that evolve by our acquaintance model with both directed and induced interactions are shown in Fig. 9. Strong macroscopic clustering emerges for intermediate values of the threshold for induced linking events.

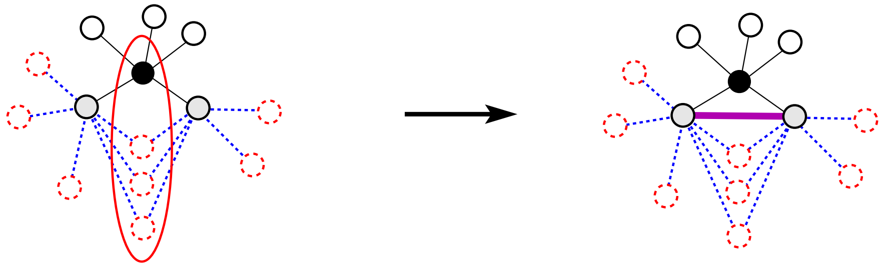
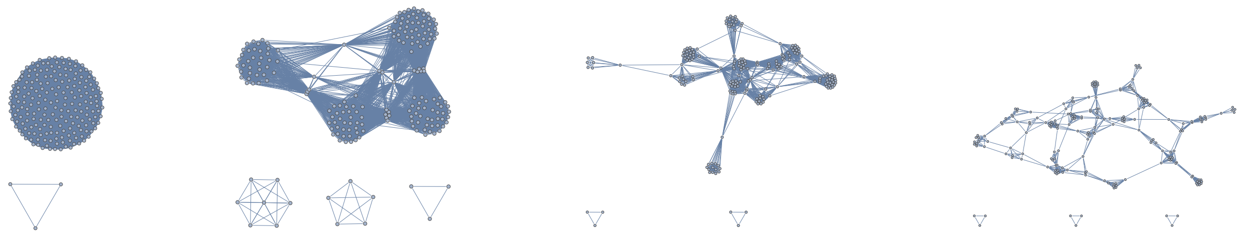


Figure 8: Illustration of induced linking in our threshold-controlled acquaintance model. An agent (solid circle) with five friends (open circles) is initially selected. This agent introduces two of them (shaded). These become friends (thick line) if the ratio of their mutual acquaintances (inside the oval) to their total acquaintances (dashed circles) exceeds a specified threshold T. Links to acquaintances external to this cluster are not shown. Here, induced linking occurs if $T < \frac{4}{9}$.



(a) (b) (c) (d)

Figure 9: Example networks of $N = 200$ for representative threshold values: (a) $T = 0.3$, (b) 0.35, (c) 0.4, and (d) 0.6. For the case $T = 0.35$ the social network breaks up into tightly-knit macroscopic communities.

Dense Networks by Node Copying

There are few generative models for dense networks *per se*. This dearth of knowlege motivated our investigation of networks that evolve by node "copying", a mechanism that generates dense networks with a host of intriguing properties. In our copying model (see also [19]), nodes are introduced sequentially and each connects to a random pre-existing target node, as well as independently to *each* neighbor of the target (friends of a friend) with probability p (Fig. 10). This simple mechanism underlies social media, such as Facebook, where an individual is invited to make a connection with a friend of a friend (see, e.g., Ref. [20]).

Specifically, a new node is connected to a randomly chosen existing node (direct linking), and it also "copies" its connections (induced linking) as indicated in Fig. 10. This copying is imperfect in that the new node links to each neighbor of a selected node in the existing network with fixed probability p.

This simple copying mechanism plays a dramatic role on the network evolution. As illustrated in Fig. 11, the network densifies as the copying parameter p is varied. When $p < 1/2$, the network is sparse, as the average number of links L grows with the total

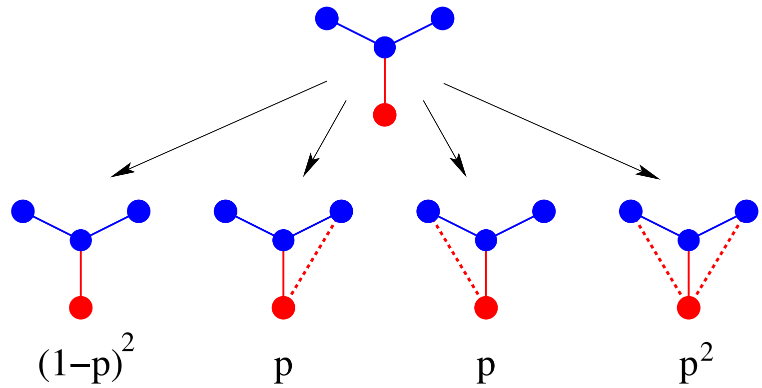


Figure 10: Illustration of the link copying model. A new agent joins a network (red). Subsequently, the each possible link to a second-order friend is independently copied with probability p.

number of nodes N as $N/(1-2p)$. In contrast, when $p > 1/2$, L grows super-linearly with the network size, namely as N^{2p}. Thus for a fixed value of $p > 1/2$, the average node degree diverges with the size of the network. Thus the network densifies as it grows and the resulting network is *non-sparse*. At the boundary between the sparse and dense regime, $p = 1/2$, the number of links grows as $N \ln N$.

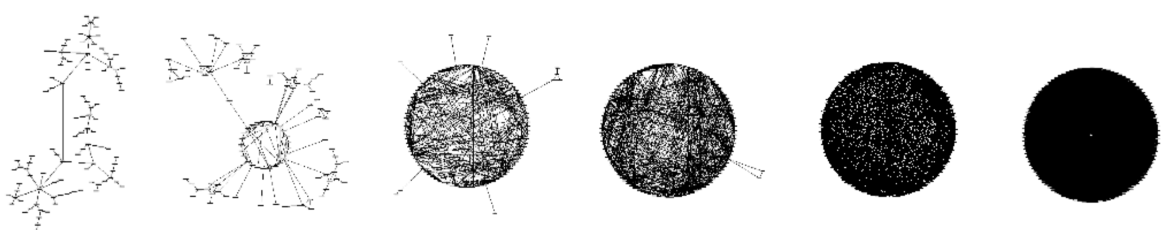


Figure 11: Illustration of the network structure in the copying model for the cases of copying probability $p = 0$, 0.2, 0.4, 0.6, 0.8, and 1 (left to right).

Based on the visual densification of the network (Fig. 11), we studied the behavior of the number of elemental motifs, such as links, triads, and higher-order cliques. To obtain the N dependence of the number of links $L(N)$, consider how the network evolves when a new node attaches to an existing node of degree k. In this case, $1+pk$ additional links are created; the factor 1 accounts for the initial link and the factor pk for the links to friends of the target. Averaging over all targets, the number ΔL of new links added is $1+p\langle k\rangle = 1+2pL/N$. The asymptotic solution to this recursion is

$$L(N) = \begin{cases} N/(1-2p) & p < \frac{1}{2}, \\ N \ln N & p = \frac{1}{2}, \\ N^{2p}/\big[(2p-1)\Gamma(1+2p)\big] & \frac{1}{2} < p \leq 1. \end{cases} \tag{13}$$

For $0 < p < \frac{1}{2}$, the network is sparse, as L is linear in N. For $p > \frac{1}{2}$, a dense network arises, as the mean degree, $2L/N$, grows with N. A related, but more intricate series of structural transitions arise for higher-order m-cliques $Q_m(N)$, namely, complete subgraphs of m nodes. For example, the number of triangles $Q_3(N) \sim L(N)$ for $p < \frac{2}{3}$, while $T \sim L^{3p/2}$ for $p > \frac{2}{3}$, and similarly for $m > 3$.

In the dense regime of $p > \frac{1}{2}$, we also found that the degree distribution now has a log-normal degree form, with strong correlations in the degrees of nearby nodes and huge fluctuations in network observables between different realizations.

References

[1] P. L. Krapivsky and S. Redner, "Organization of Growing Random Networks", Phys. Rev. E **63**, 066123 (2001).

[2] A. Gabel and S. Redner, J. Stat. Mech. P02043 (2013).

[3] A. Gabel, P. L. Krapivsky, and S. Redner, Phys. Rev. E. **88**, 050802 (2013).

[4] A. Gabel, P. L. Krapivsky, and S. Redner, arXiv:1312.7843.

[5] P. L. Krapivsky and D. Krioukov, Phys. Rev. E **78**, 026114 (2008).

[6] R. F. i Cancho and R. V. Solé, *Statistical Mechanics of Complex Networks*, no. 625 in Lecture Notes in Physics, p. 114, Springer, Berlin (2003).

[7] G. Bianconi and A.-L. Barábasi, Phys. Rev. Lett. **86**, 5632 (2001); G. Bianconi and A.-L. Barábasi, Europhys. Lett. **54**, 436 (2001).

[8] P. L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002).

[9] D. L. Bryan and M. E. O'Kelly, J. Regional Science, **39**, no. 2, p. 275 (1999).

[10] J. J. Han, N. Bertain, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupay, A. J. M. Walhout, M. E. Cusick, F. P. Roth, and M. Vidal, Nature **430**, 88 (2004).

[11] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **102**, 7794 (2005).

[12] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[13] J. Kunegis, M. Blattner, and C. Moser, arXiv:1303.6271.

[14] P. L. Krapivsky and S. Redner, J. Phys. A **35**, 9517 (2002).

[15] P. L. Krapivsky and S. Redner, J. Stat. Mech. P06002 (2013).

[16] R. M. D'Souza, P. L. Krapivsky, and C. Moore, Eur. Phys. J. B **59**, 535 (2007).

[17] P. L. Krapivsky and S. Redner, J. Stat. Mech. P04021 (2014).

[18] U. Bhat, P. L. Krapivsky, and S. Redner, J. Stat. Mech. P11035 (2014).

[19] P. L. Krapivsky and S. Redner, Phys. Rev. E. **71**, 036118 (2005).

[20] M. O. Jackson and B. W. Rogers, Amer. Econ. Rev. **97**, 890 (2007).

[21] P. L. Krapivsky and S. Redner, "Organization of Growing Random Networks", Phys. Rev. E **63**, 066123 (2001).

[22] H. Rozenfeld and D. ben-Avraham, "Designer Nets from Local Strategies", Phys. Rev. E **70**, 056107 (2004).

[23] E. Ben-Naim and P. L. Krapivsky, "Random ancestor trees", J. Stat. Mech. P06004 (2010).

[24] M. E. J. Newman, *Networks: An Introduction* (Oxford Univ. Press, Oxford, 2010).

[25] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, 1953).

[26] P. Erdős and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).

[27] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

[28] S. Janson, T. Łuczak, and A. Rucinski, *Random Graphs* (John Wiley & Sons, New York, 2000).

[29] K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, 1991).

[30] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, New York, 1994).

[31] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2000).

[32] P. L. Krapivsky and S. Redner, "A Statistical Physics Perspective on Web Growth", Computer Networks **39**, 261 (2002).

1.

1. Report Type

Final Report

Primary Contact E-mail
Contact email if there is a problem with the report.

redner@santafe.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

+1-505-946-2764

Organization / Institution name

Boston University

Grant/Contract Title
The full title of the funded effort.

Network Dynamics: Modeling and Generation of Very Large Heterogeneous Social Networks

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0391

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Sidney Redner and Paul L. Krapivsky

Program Manager
The AFOSR Program Manager currently assigned to the award

Reza Ghanadan

Reporting Period Start Date

08/01/2012

Reporting Period End Date

7/31/2015

Abstract

One major achievement was the construction of redirection algorithms to efficiently generate large networks with prescribed degree characteristics. A hindered redirection algorithm was shown to reproduce sublinear preferential attachment. Conversely, enhanced redirection leads to highly-dispersed networks that contain multiple macrohubs (degree a finite fraction of the number of network nodes) and exhibit non-extensive scaling.

The average number of distinct degrees that appear in a finite network was found to grow algebraically with network size and the underlying distribution is a universal Gaussian. A choice-driven network growth mechanism was formulated in which a new node first identifies a set of target nodes and attaches to either the target with the largest degree (greedy choice), or the target whose degree is not the largest (meek choice). The resulting network exhibits a non-universal power-law degree distribution.

An algorithm to generate heterogeneous social networks from homogeneous interactions was constructed. Heterogeneity emerges dynamically, rather than being imposed explicitly by homophily or by heterogeneous agents. Finally, a node copying mechanism was formulated to generate dense networks. These networks exhibit multiple structural phase transitions in the densities of cliques (small complete

graphs) and a log-normal degree distribution in the dense regime.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

2015 Redner (AFOSR-FA9550-12-1-0391) Form 298.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

2015-Redner (AFOSR-FA9550-12-1-0391) final-rept.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

See attached report

Changes in research objectives (if any):

N/A

Change in AFOSR Program Manager, if any:

Prior to Reza Ghanadan, the Program Manager was Tony Falcone

Extensions granted or milestones slipped, if any:

N/A

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Nov 10, 2015 12:55:16 Success: Email Sent to: redner@santafe.edu