Networking the seceder model: Group formation in social and economic systems

Andreas Grönlund* and Petter Holme†
Department of Physics, Umeå University, 901 87 Umeå, Sweden
(Received 29 November 2003; revised manuscript received 12 May 2004; published 16 September 2004)

The seceder model illustrates how the desire to be different from the average can lead to formation of groups in a population. We turn the original, agent based, seceder model into a model of network evolution. We find that the structural characteristics of our model closely match empirical social networks. Statistics for the dynamics of group formation are also given. Extensions of the model to networks of companies are also discussed.

DOI: 10.1103/PhysRevE.70.036108 PACS number(s): 89.65.−s, 89.75.Hc

I. INTRODUCTION

Social networks have "community structure"—actors (vertices) with the same interests, profession, age (and so on), organize into tightly connected subnetworks, or communities [1–3]. Subnetworks are connected into larger conglomerates in a hierarchical structure of larger and more loosely connected structures. Over the last few years the issue of communities in social networks has ventured beyond sociology into the area of physicists' network studies [4–6]. The problem of how to detect and quantify community structure in networks has been the topic of a number of papers [2,7,8], whereas a few others have been models of networks with community structure [9–11]. In these models, the common properties defining the community are external to the network evolution (in the sense that an individual does not choose the community to belong to by virtue of his or her position in the network). In this paper we present a model where the community structure emerges as an effect of the agents personal rationales. We do this by constructing a networked version of an agent based model—the seceder model [12–15]—of social group formation based on the assumption that people actively try to be different than the average. Independence and the desire to be different play an important role in social group formation [16], this might be even more important in the social networking of adolescents. The important observation is that few want to be different from *anyone* else, rather one tries to affiliate to noncentral group. This type of mechanism is probably rather ubiquitous, so the connotations of eccentricity are not intended for the name of the model. (See Ref. [17] for a nonscientific account of the formation of youth subcultures by these and similar premises.)

Another system where the networked seceder model can serve as a model—or at least a direction for extension of present models (see, e.g., Ref. [18])—is networks where the vertices are companies and the edges indicate a similar niche. (Such edges can be defined indirectly using stock-price correlations [19].) The establishment of new companies are naturally more frequent in new markets. Assuming new markets are remote to more traditional markets, the networked seceder model makes a good model of such company networks.

II. PRELIMINARIES

A. Notations

The model we present produces a sequence (or time evolution) of graphs $\{G_t\}$. Each graph in this sequence consists of the same set V of N vertices, and a time specific set of M undirected edges E_t. The model defines a Markov process and is thus suitable for a Monte Carlo simulation. The number of iterations of the algorithm defines the simulation time $t=1,\ldots,t_{max}$.

We let $d(i,j)$ denote the distance (number of edges in the shortest path) between two vertices i and j. We will also need the *eccentricity* defined as the maximal distance from i to any other vertex.

B. The seceder model

The original seceder model [12] is based on N individuals with a real number $s(i)$ representing the traits (or personality) of individual i. The algorithm is then to repeat the following steps.

(1) Select three individuals i_1, i_2, and i_3 with uniform randomness.

(2) Pick the one (we call it $\hat{i}$) of these whose s-value is farthest away from the average $[s(i_1)+s(i_2)+s(i_3)]/3$.

(3) Replace the s value of a uniformly randomly chosen agent j with $s(\hat{i})+\eta$, where η is a random number from the normal distribution with mean zero and variance one.

Note that the actual values of s are irrelevant, only the differences between s of different agents. The output of the seceder model is a complex pattern of individuals that stick together in well-defined groups. The groups have a life cycle of their own—they are born, spawn new groups, and die. Statistical properties of the model are investigated in Ref. [12], effects of a bounded trait space is studied in Ref. [13], the fitness landscape is the issue of Ref. [14] and Ref. [15] presents a generalization to higher-dimensional trait spaces.

Our generalization of this model to a network model is based on the idea that if the system is embedded in a net-

*Electronic address: gronlund@tp.umu.se
†Electronic address: holme@tp.umu.se

work, then the difference in personality is implicitly expressed through the network position. Thus, the identity number (or vector) s becomes superfluous in a network model. The homophily assumption [20]—that like attracts like—means that the difference in character between two vertices i and j (defined as $|s(i)-s(j)|$ in the traditional seceder model) can be estimated by the graph distance $d(i,j)$ in a networked model. The model we propose is then, starting from any graph with N vertices and M edges, to iterate the following steps.

(1) Select three different vertices i_1, i_2, and i_3 with uniform randomness.

(2) Pick the one $\hat{i}$ of these that is least central in the following sense: If the graph is connected, vertices of highest eccentricity are the least central. If the graph is disconnected the most eccentric vertices within the smallest connected subgraph are the least central. If more than one vertex is least central, let $\hat{i}$ be a uniformly randomly chosen vertex in the set of least central vertices.

(3) Select another vertex j in $V\backslash\{\hat{i}\}$ uniformly randomly. If $\deg j < \deg \hat{i}+1$, rewire all of j's edges to $\hat{i}$ and a random selection of $\hat{i}$'s neighbors. (By rewire an edge (v,w) of a vertex v we mean that (v,w) is replaced by (v,w'), $v \neq w$, in E.) If $\deg j \geqslant \deg \hat{i}+1$, rewire j's edges to $\hat{i}$, $\hat{i}$'s neighborhood vertices and (if $\deg j > \deg \hat{i}+1$) to $\deg j - \deg \hat{i} - 1$ randomly selected other vertices.

(4) Go through all j's edges once more and, with a probability p, rewire these.

The rewiring of steps 3 and 4 are performed with the restriction that no multiple edges or loops (edges that go from a vertex to itself) are allowed. Steps 1–3 correspond rather closely to the same steps of the original model. That j's edges are rewired mainly to the neighborhood of $\hat{i}$ (and $\hat{i}$ itself) reflect the inheritance of trait value in the original model—by the homophily assumption, the neighborhood of $\hat{i}$ will have much the same traits as $\hat{i}$. The main difference between the original and the networked seceder model is step 4 where some edges are rewired to distant vertices. The motivation for this step is that long-range connections exist in real-world networks [21,22], and can in some situations be even more important than the strong links within a group [23]. This kind of rewiring, to obtain long-range connections has been used to model "small-world behavior" of networks [21] (i.e., a logarithmic, or slower, scaling of the average intervertex distance for ensembles of graphs with a constant average degree [4]).

To make the model consistent we also have to specify the initial graph. As far as we can see, at least for finite p, this choice is irrelevant—the structure of the generated graphs is the same (or at least very similar). We will not investigate this point further. Instead we fix the initial graph to an instant of Erdős and Rényi's random graph model [24] (for a modern survey of this model, see Ref. [25]): A graph with N edges and M edges is constructed by starting from isolated vertices and then iteratively introduce edges between vertex pairs chosen by uniform randomness and with the restriction that no multiple edges or loops are allowed. To be sure that the structure of the random graph is gone we run the construction algorithm $10N$ sweeps through every vertex before the graph is sampled. (We justify this number *a posteriori* later.)

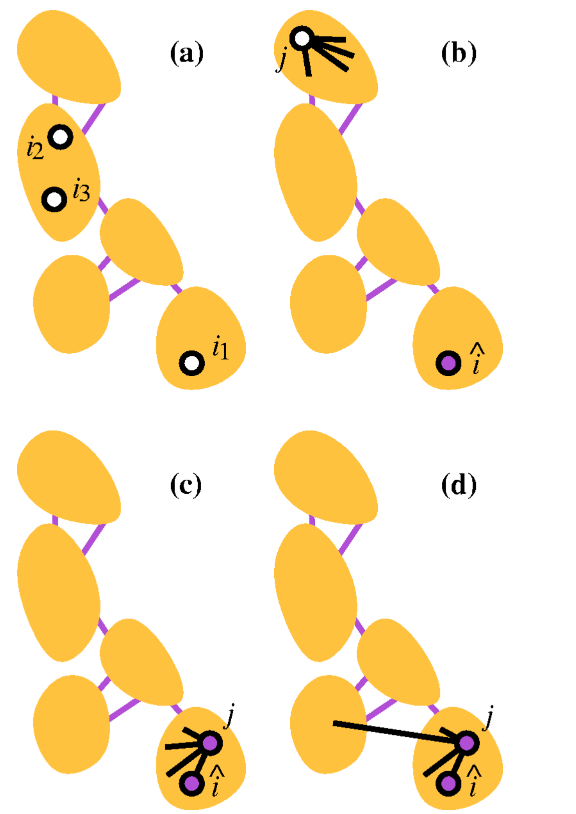

FIG. 1. Illustration of the networked seceder model. (a) In step 1 three vertices, i_1, i_2, and i_3, are chosen at random. (b) In step 2 the least central of the three vertices is relabeled to $\hat{i}$. In step 3 a vertex j is selected at random and (c) the edges of j are rewired to $\hat{i}$ and $\hat{i}$s neighborhood (and to a set of random other vertices if necessary). Note that, in (c), j is moved to the cluster it is rewired to. In step 4 j's edges are rewired with a probability p. The shaded areas represent tightly connected subgraphs.

An illustration of the construction algorithm can be seen in Fig. 1. A realization of the algorithm is displayed in Fig. 2. The p value of this realization is zero. For the value $p=0.1$ we use in most simulations the community structure is less visible to the eye. Nevertheless—as we will see—the community structure is still substantial for much larger values of p.

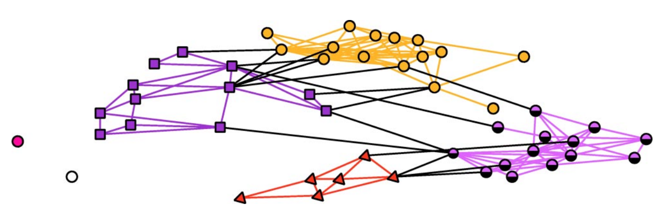

FIG. 2. One realization of the networked seceder model. The model parameters are $N=50$, $M=150$, and $p=0$. The indicated groups are indentified with Newman's clustering algorithm (see Sec. II C). This realization has modularity $Q=0.575$, clustering coefficient $C=0.530$, and assortative mixing coefficient $r=0.0456$.

C. Detecting communities

To analyze the structure of cohesive subgroups in our model networks we use the community detection scheme presented in Ref. [26]. This algorithm starts from one-vertex clusters and (somewhat reminiscent of the algorithm in Ref. [27]) iteratively merge clusters to form clusters of increasing size with relatively few edges to the outside. The crucial ingredient of this scheme is a quality function

$$Q' = \sum_{s\in S} (e_{ss} - a_s^2), \qquad (1)$$

where S is the set of subnetworks at a specific iteration of the algorithm and $e_{ss'}$ is the fraction of edges that goes between a vertex in s and a vertex in s', and $a_s = \Sigma_{s'} e_{ss'}$. The algorithm performs a steepest accent in Q' space—at each iteration the two clusters that give the largest increase (or smallest decrease) of Q' are merged. The iteration having the highest Q' value—which defines the modularity Q—gives the partition into subgroups.

D. Conditional uniform graph tests

One can argue that some network structures are more basic than others. Given such an assumption and a network G, an interesting issue is whether a certain structure, say X, is an artifact of a more basic structure, say Y. One way to do this is by a conditional uniform graph test: One compares the value of $X(G)$ with X averaged over an ensemble of graphs with the value of Y fixed to $Y(G)$. This has (since Ref. [28]) been a well established technique in social network analysis and has recently been brought over to physicists' [29] and biologists' [30] network literature. A common assumption [29–31] is that the degree distribution is such a very basic structure. We make this assumption too and perform a conditional uniform graph test with respect to the degree sequence of the networks. To sample networks with a given degree sequence we use the idea of Ref. [31] to rewire the edges of the network in such a way that the degree sequence remains unaltered. More precisely we go through all edges $(i,j)\in E$ and perform the following.

(1) Construct the set E' of edges such that if $(\hat{i},\hat{j})\in E'$ then replacing (i,j) and $(\hat{i},\hat{j})$ by $(i,\hat{j})$ and $(\hat{i},j)$ would not introduce any loops (self-edges) or multiple edges.

(2) Pick an edge $(\hat{i},\hat{j})\in E'$ by uniform randomness.

(3) Rewire (i,j) to $(i,\hat{j})$ and $(\hat{i},\hat{j})$ to $(\hat{i},j)$.

For every realization of the seceder algorithm we sample n_{sample}=ten randomized reference networks as described earlier. The motivation for this rather low number is that all quantities seem to be self-averaging (the fluctuations decrease with N) and many have symmetric distributions with respect to rewirings (which implies that many realization averages compensate for few rewiring averages). To further motivate this small n_{sample} we compare with $n_{\text{sample}}=100$ for the smallest size ($N=200$, which, as mentioned, is most affected by fluctuations) and find that the quantities typically differ by 0.5% which we consider small.

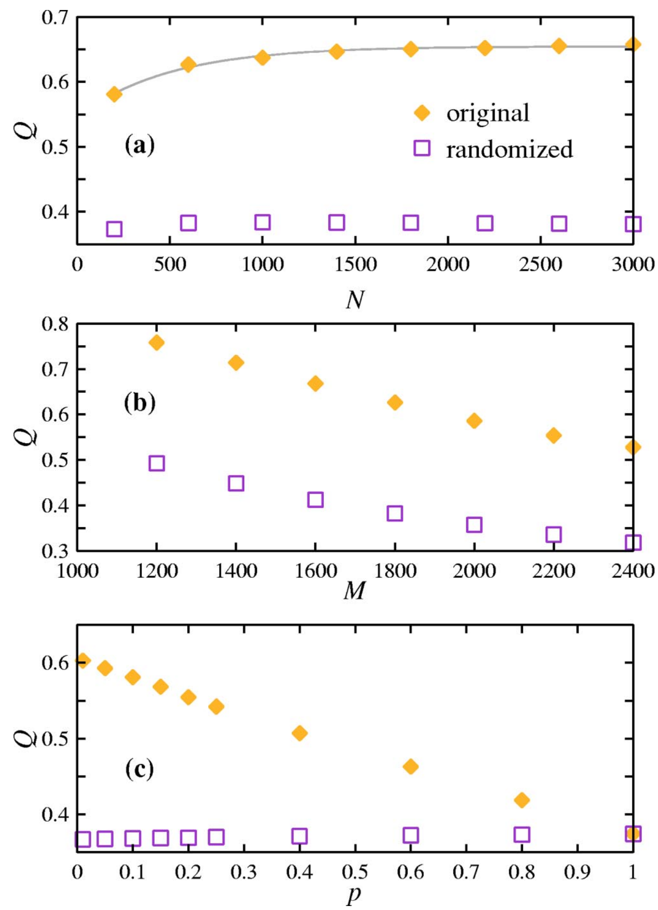

FIG. 3. The modularity Q as a function of the model parameters. (a) shows Q as a function of N with $M=3N$ and $p=0.1$. (b) displays Q for different M for $N=600$ and $p=0.1$. In (c) we plot the p dependence of Q for $N=200$ and $M=600$. The gray line in (a) is a fit to an exponential. All error bars are smaller than the symbol size.

III. THE COMMUNITY STRUCTURE OF THE SECEDER MODEL

The key quantity capturing the degree of community order in the network is the modularity Q (defined in Sec. II C). In Fig. 3(a) we see that, if the average degree and p is kept constant then Q converges to a high value, $Q\approx 0.64$ for $p=0.1$ and $M=3N$. This value is much higher than the reference value from the randomized networks—this curve has a peak around $N=1500$ and decays for larger N, larger sizes would be needed to see if Q converges to a finite value for the randomized networks. With the analogy to the Watts-Strogatz model (where a fraction p of a circulant's [32] edges is rewired randomly) we would say that $p=0.1$ is a rather high value, still Q is much higher for the networked seceder model than for random networks with the same degree distribution. From this we conclude that our model fulfills its purpose—it produces networks with a pronounced community structure just as the original seceder model makes agents divide into well-defined groups in trait space. In Fig. 3(b) we plot the M dependence of Q for fixed $N=600$ and $p=0.1$. We see that Q decreases with M for both the seceder model and the randomized networks. As M approaches its maximum

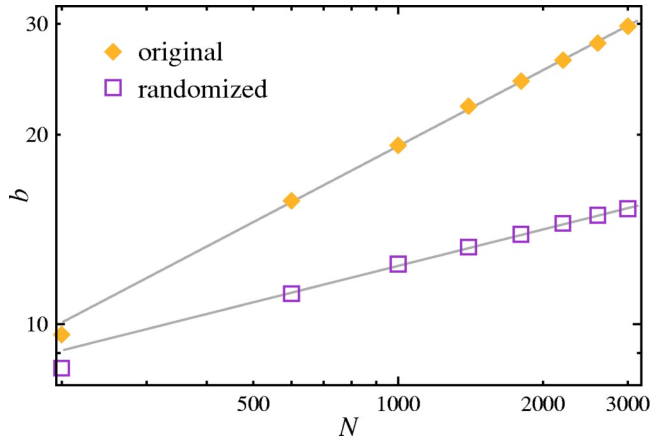

FIG. 4. The number of groups b as a function of the system size N. The other parameter values are $M=3N$ and $p=0.1$. The line is a fit to a power-law ab^{β}. For this set of parameters $\beta=0.400(6)$ for the seceder model and 0.193(6) for the reference graphs of the conditional uniform graph test. All error bars are smaller than the symbol size. Note the double-logarithmic scale.

value $N(N-1)/2$ the curves will converge (since the fully connected graph is unique), but the figure shows that the curves are separated for a wide parameter range. More importantly it suggests that the quantity Q should be rescaled by some appropriate function if networks of different average degree are to be compared. In the rest of our paper, however, we will keep the degree constant. In Fig. 3(c) we show the p dependence of Q. As expected Q decays monotonously, in fact almost linearly, with p. The curves for the seceder model converge to the curve of the randomized networks as $p\to 0$. Q of the randomized reference networks is almost p independent. The fact that it is not completely p independent means that the degree distribution of the seceder model must vary with p. We will strengthen this claim later.

Figure 4 shows the size dependence of b—the number of groups. We see that this function can be well described by a constant plus a power law

$$A + b^{\beta} \tag{2}$$

(where A is a constant) with an exponent $\beta=0.400(6)$ for the seceder model and $\beta=0.193(6)$ for the random networks with the same degree distribution. The average community size is given by N/b and will therefore also behave as a power law, with exponent $1-\beta=0.600(6)$. This fact—that the number and average size of the communities grow with N—does not seem contradictory to the real world to us. Since a community, both in a social and economical interpretation of the model, does not need to be controlled or supervised there is no natural upper limit to the number of community members. Furthermore, there is no particular constraint on the number of communities present in real world systems. A thorough study of the scaling exponents would be interesting, but falls out of the scope of the present paper.

In Fig. 5 we display the average geodesic lengths within a community l_{intra} and between vertices of different communities l_{inter} for parameter values $M=3N$ and $p=0.1$. To be precise, we consider the largest connected component (which typically contains 99% of the vertices), and define

$$l_{\text{intra}} = \frac{1}{N_{\text{intra}}} \sum_{i=1}^{b} \sum_{v,w\in B^i} d(v,w) \tag{3a}$$

and

$$l_{\text{inter}} = \frac{1}{\binom{N}{2} - N_{\text{intra}}} \sum_{i=1}^{b} \sum_{v\in B^i} \sum_{w\notin B^i} d(v,w) \tag{3b}$$

where B^i is the ith cluster and

$$N_{\text{intra}} = \sum_{i=1}^{b} \binom{i}{2} \tag{4}$$

is the number of pairs of vertices belonging to the same community. As seen in Figs. 5(a) and 5(b) both l_{intra} and l_{inter}

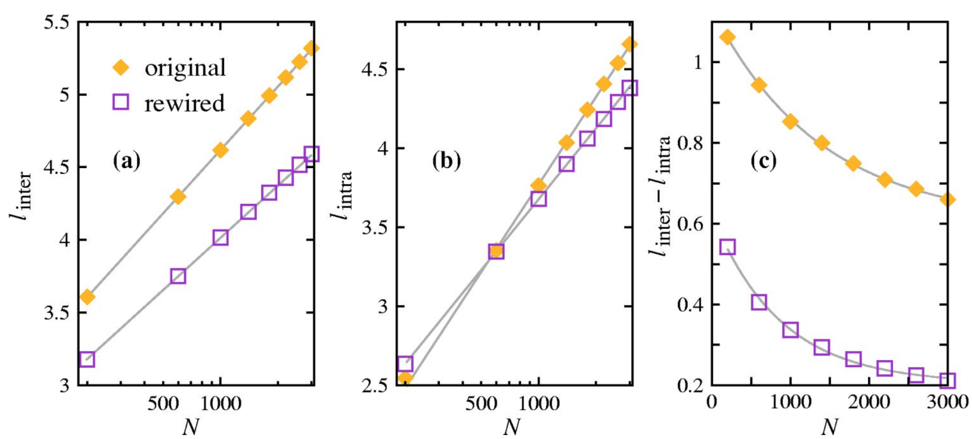

FIG. 5. Average distance between, and within, clusters (as identified by the algorithm described in Sec. II C). The gray lines are fits to an exponential form. The slope of the original is the same in (a) and (b) [also the rewired line has the same slope in (a) and (b)]. All error bars are smaller than the symbol size.

grow logarithmically as functions of N with the same slope in a semilogarithmic plot. A logarithmic scaling of the average shortest path length (which of course also holds) is expected (cf. Ref. [33]). But we could not anticipate the lack of qualitative difference between distances between vertices of the same and different clusters. The actual values of l_{intra} is significantly smaller than l_{inter} and this difference holds as $N\to\infty$: As seen in Fig. 5(c) $l_{\text{inter}}-l_{\text{intra}}$ converges to 0.60(1). The same value for the randomized graphs is $l_{\text{inter}}-l_{\text{intra}}=0.204(8)$ which is expected—the detected communities in the networked seceder model are more well defined and tight-knit than the corresponding communities in a random network with the same degree distribution.

IV. OTHER STRUCTURAL CHARACTERISTICS

Apart from the quantities of the previous section (all directly related to the community structure), we also look at some other well established structural measures: The degree distribution, the clustering coefficient, and the assortative mixing coefficient.

A. Degree distribution

Following the works of Barabási and co-workers [34–36] the degree distribution has been perhaps the most studied network structure. In some social networks—of telephone calls [37], e-mail communication [38], and the network of sexual contacts [39]—the degree distribution fits well to a power-law functional form. Other social network studies report right skewed degree distributions that deviate from a power law in either the high- or low-k limit [40–43]. Yet other studies have found social networks with Gaussian degree distributions [40,44,45], or exponential degree distributions [3,46]. We conclude that the degree distribution of social networks still is an open question with, most likely, not a single solution—different social networks may follow different degree distributions. The degree distribution of the networked seceder model is displayed in Fig. 6. We note that $P(k)$ has an exponential tail, notably larger than the Poisson degree distribution. Clearly this falls into one of the cases mentioned earlier.

B. Clustering coefficient

The clustering coefficient C measures the fraction of connected triples of vertices that form a triad. This type of statistics has been popular since Ref. [21]. The definition we use is slightly different from that of Ref. [21]:

$$C=\frac{c(3)}{p(3)}, \tag{5}$$

where $c(n)$ denotes the number of representations of circuits of length n and $p(n)$ denotes the number of representations of paths of length n. (By “representation” we mean an ordered triple such that one vertex is adjacent to the vertex before or after. For example, a triangle has six representations—all permutations of the three vertices.) This definition is common in sociology (although sociologists em-

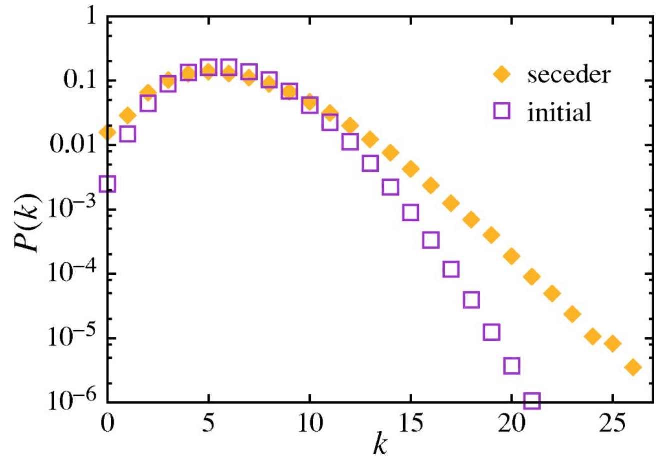

FIG. 6. Degree distribution of the networked seceder model. The model parameters are $N=1800$, $M=5400$, and $p=0.1$. The squares indicate the degree distribution of a random graph with the sizes (N and M), i.e., the initial network before the iterations of the seceder model commence.

phasize triad statistics for directed networks)—see Ref. [47] for a review—but is also frequent in physicists' literature since Ref. [48]. A plot of C as a function of N is shown in Fig. 7(a). We see that C for the seceder model converges to a constant value rather rapidly. Similarly the C for the rewired networks goes to zero roughly over the same time scale. The fact that community structure induces a high clustering is well known and modeled [49], as is the fact that the clustering vanishes like $1/N$ in a random graph with Poisson degree distribution [4].

In Fig. 7(b) we plot the local clustering coefficient

$$C_v=\frac{|\Gamma_v|_E}{\binom{k}{2}} \tag{6}$$

as a function of the degree k of the vertex ($|H|_E$ denotes the number of edges in a subgraph H, Γ_v is the neighborhood of v). C_v [21] measures how well connected the neighborhood of v is—if $C_v=0$ none of the vertices in v's neighborhood has an edge to any other, if $C_v=1$ there is an edge between each pair of vertices in v's neighborhood. We find that for the seceder model the local clustering coefficient is roughly inversely proportional to the degree. For the rewired reference network, on the other hand, C_v is independent of the degree. It is known that many real-world networks, including social networks, show the same $C_v\sim k^{-1}$ scaling as the seceder model [50]. It is furthermore known that some simple network models, like the Barabási-Albert model [34], has a k-independent local clustering—just as the randomized reference networks [50].

C. Degree-degree correlations

The assortative mixing coefficient [51] is the Pearson correlation coefficient of the degrees at either side of an edge

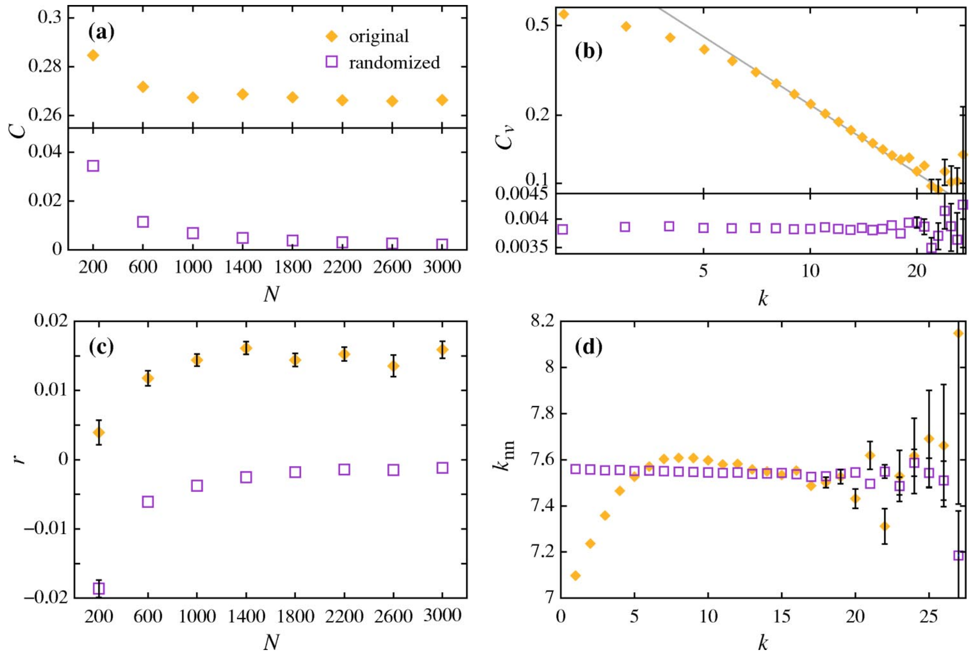

FIG. 7. Common structural measures. (a) shows the clustering coefficient C as a function of the number of vertices for the seceder model and rewired networks. (b) shows the local clustering coefficient C_v as a function of the degree k of v. The gray line is inversely proportional to k. (Note the double logarithmic scale.) (c) displays the corresponding plot of the assortative mixing coefficient. (d) shows the average degree of the neighbors of a vertex as a function of the vertex's degree. The network parameters are $M=3N$ and $p=0.1$, in (b) and (d) $N=1800$. Error bars are shown if they are larger than the symbol size.

$$r = \frac{4\langle k_1 k_2\rangle - \langle k_1 + k_2\rangle^2}{2\langle k_1^2 + k_2^2\rangle - \langle k_1 + k_2\rangle^2}, \tag{7}$$

where subscript i denotes the ith argument and averaging is over the edge set. r is known to be positive in many social networks [51,52]. It has been suggested that this assortative mixing can be related to community structure [53]. Against this backdrop it is not surprising to note that the networked seceder model produces networks with markedly positive r, see Fig. 7(c). The reference networks with the same degree sequences converge to zero from negative values, as also observed in Ref. [42]. It has been argued [29,54] that networks formed by agents without any preference for the degrees of the neighboring vertices gets negative r from the restriction that only one edge can go between one pair of vertices. This is probably the reason for the negative r values of the rewired networks.

In Fig. 7(d) we give a more detailed picture of the degree-degree correlations, we plot the average neighbor degree

$$k_{nn}(v) = \frac{1}{k_v}\sum_{w\in\Gamma_v} k_w \tag{8}$$

against the degree [55]. We see that the dissortativity mainly stems from that the vertices of low degree tends to connect to other vertices of low degree. For vertices of mid and high degree, $k_{nn}(k)$ is almost independent of k.

V. CHARACTERISTICS OF COMMUNITY DYNAMICS

In this section we look at the dynamics of the communities. To do this we need criteria for if a cluster B_k^t at time t is the same as cluster $B_{k'}^{t-1}$ at time $t-1$. The idea is to find the best possible matching of vertices between the partition into clusters of the two consecutive time steps (see Fig. 8). To give a mathematical definition, let $\mathcal{B}_t=\{B_t^1,\ldots,B_t^{b(t)}\}$ be the partition of G_t into clusters by the algorithm described in Sec. II C and let $b'=\min[b(t),b(t-1)]$. We define a mapping

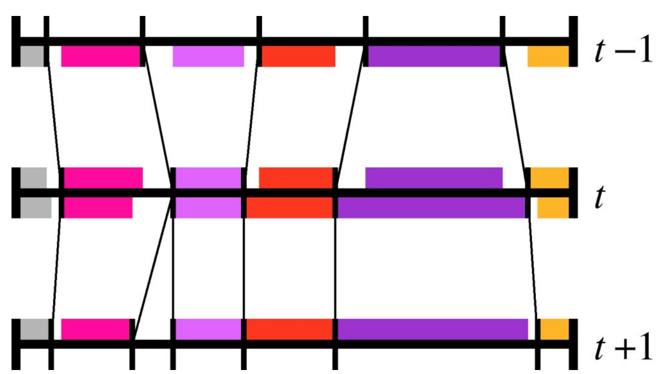

FIG. 8. Illustration of the identification of clusters at consecutive time steps. The vertex set is represented by the horizontal line. The vertical tics demarcate cluster boundaries. The communities at consecutive time step are matched so that the overlap (the horizontal sum of shaded segments) is maximized.

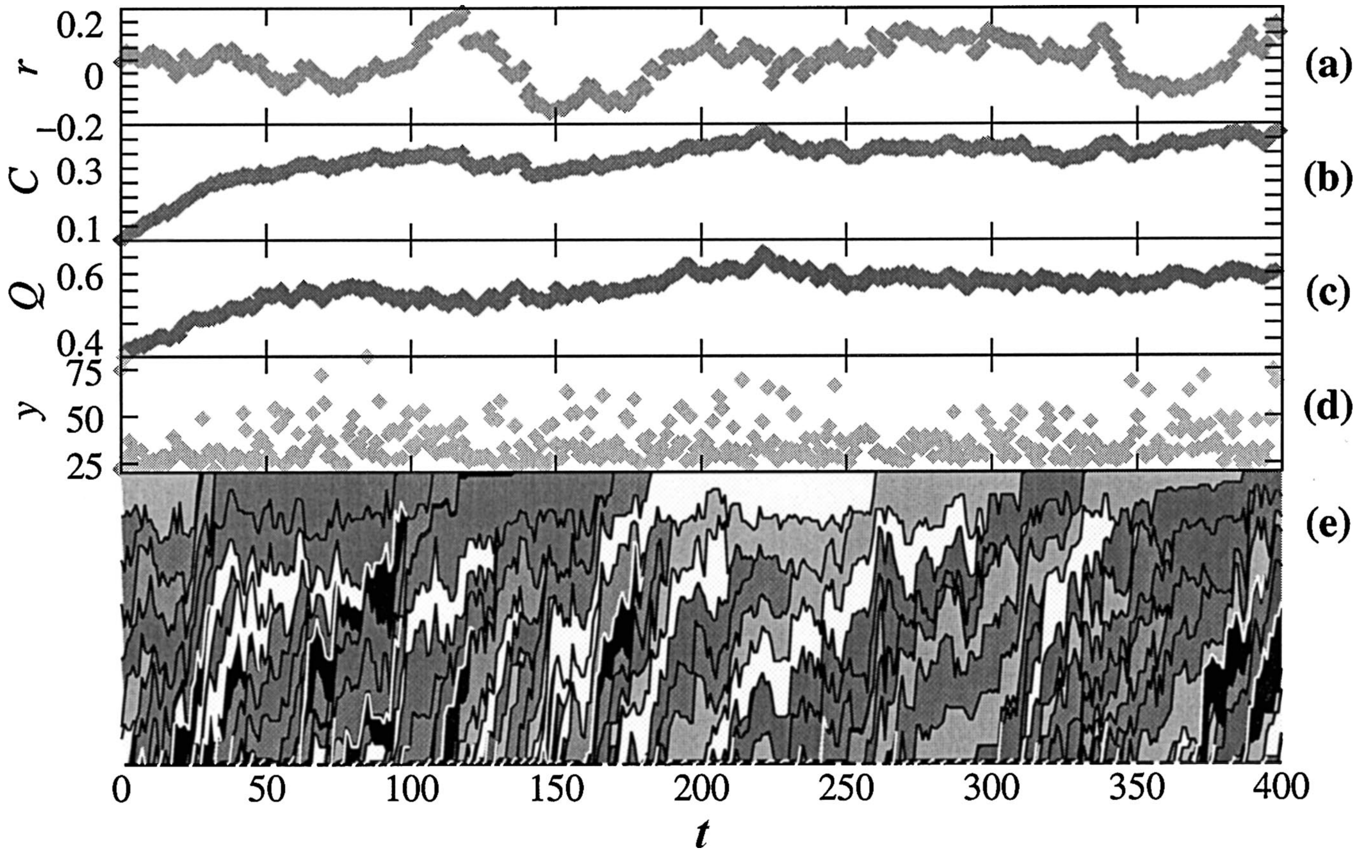

FIG. 9. Community dynamics for a typical run with the parameter values $N=100$, $M=300$, and $p=0$ and 100 iterations of the networked seceder model. The different panels show different statistics for one single run of the algorithm. (a) shows the time evolution of the assortative mixing coefficient. (b) shows the clustering coefficient C. (c) shows the modularity Q. (d) shows the maximal overlap y between consecutive time steps. (e) illustrates the time evolution of the communities. A vertical cross section of (e) gives the respective relative sizes of the different clusters. The clusters are sorted horizontally according to age—the oldest clusters are in the top of the panel.

f from b' elements of $[1,b(t-1)]$ to b' elements of $[1,b(t)]$ such that the overlap

$$y'_t = \sum_{k=1}^{b'} |B^k_{t-1} \cap B^{f(k)}_t| \tag{9}$$

is maximized ($|\cdot|$ denotes cardinality). Let $y(t)$ denote this maximized y'_t value. To calculate this overlap we use the straightforward method of testing all matchings. In principle, this algorithm runs in exponential time, but since the number of groups is typically rather low, systems of a few hundred vertices are numerically tractable.

The evolution of the group structure, with the group structure identified as described above, is displayed in Fig. 9. In Figs. 9(a) and 9(b) we see the time evolution of the assortative mixing coefficient r and the clustering coefficient C, whose average size scaling was studied in Sec. IV. We note that the assortative mixing coefficient fluctuates rather much. Even though it is mostly positive (remember that the average value is significantly positive) it can also be negative. This is likely to be a finite size phenomenon—as the assortative mixing coefficient is self-averaging [42], larger systems would not fluctuate much and have stable positive values [as seen in Fig. 7(b)]. The clustering coefficient as displayed in Fig. 9(b) shows a more stable evolutionary trajectory. Over a time scale roughly corresponding to $N=100$ updating steps C goes from the value of the initial Erdős-Rényi graphs to the higher clustering coefficient of the networked seceder model. This is natural since it is also roughly the time scale for all vertices to be picked and rewired once. The value of the modularity Q, displayed in Fig. 9(c), shows a similar behavior as the clustering coefficient as it increases from the value ~ 0.4 of the original random graph to ~ 0.6 of the seceder model. C and Q seem to be strongly correlated, something that seems very logical in the context of the seceder model—the clustering coefficient increases when a high degree vertex is rewired to a specific cluster, a process that also strengthens the community structure. If this strong C-Q correlation is a ubiquitous property it is an interesting problem for future studies. In Fig. 9(d) we plot the overlap y which fluctuates between 25 and 75 with an average well below 50. These values are lower than we expected *a priori*, as it means that identity of more than half the group members change at a typical time step. Just as the fluctuations in r, we expect the fluctuations in the cluster structure to decrease with system size, therefore y/N will increase with N. In Fig. 9(e) the time development of different cluster sizes is illustrated. A horizontal cross section gives the size partitioning of the vertex set at a given time step. A demarcated area represents a group. Older groups are above younger groups. An observation from Fig. 9(e) is that groups typically live between one and 100 time steps. The lifetime scale of groups seems to coincide with that of the initial relaxation to the seceder equilibrium. We also note that there seems to be no particular correlation between age and stability or size, a situation that would have produced skewed lifetime or cluster-size distributions.

The observations in this section were checked for a few other runs and seem to be representative. Since they do not hint some surprising phenomena we do not conduct any extensive statistical survey of the dynamical properties.

VI. SUMMARY AND CONCLUSIONS

We have proposed a model for network formation based on the seceder model. The model captures how a community structure can emerge from the desire to be different, both in social and economic systems. The community structure of our model is analyzed with a recent graph clustering scheme. This scheme has the advantage that it gives a measure of the degree of community structure in a network—the modularity Q. We see that the Q is much higher for our model networks than for random reference networks with the same degree distributions. Both the number of groups and the average size of groups grow as power laws with sublinear exponents. Both the average geodesic distance between vertices of the same and different clusters grow logarithmically; the difference between these, however, is much larger for the networked seceder model than for the random reference networks. The general picture is thus that the networked seceder model generates well-defined communities just like the agents of the original seceder model gets clustered in trait space.

The networked seceder model gives networks of high clustering and positive assortative mixing by degree—properties that are known to be characteristic of acquaintance networks. The degree distribution has a peak around the average degree and exponentially decaying—also that consistent with real world observations.

The dynamics of the communities was briefly investigated by defining a mapping between consecutive time steps that maximizes an overlap function. Using this method we conclude that the speed of the dynamics is set by the size of the system. We see that the clustering coefficient and modularity are strongly correlated and that older groups are not necessarily larger than younger.

To epitomize, the networked seceder model gives a mechanism of emergent community structure that is different from earlier proposed mechanisms in network models [9–11]. The mechanism is arguably present in, at least, social networks [16]. We speculate that this model can be applied to networks of companies that are linked if they are active in the same market.

ACKOWLEDGMENTS

Thanks are due to Beom Jun Kim, Fredrik Liljeros, Petter Minnhagen, and Mark Newman. The authors were partially supported by the Swedish Research Council through Contract No. 2002-4135.

[1] P. Gleiser and L. Danon, Adv. Complex Syst. **6**, 565 (2003).
[2] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[3] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[6] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[7] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[8] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).
[9] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003).
[10] E. M. Jin, M. Girvan, and M. E. J. Newman, Phys. Rev. E **64**, 046132 (2001).
[11] B. Skyrms and R. Freemantle, Proc. Natl. Acad. Sci. U.S.A. **97**, 9340 (2000).
[12] P. Dittrich, F. Liljeros, A. Soulier, and W. Banzhaf, Phys. Rev. Lett. **84**, 3205 (2000).
[13] P. Dittrich, InterJournal 363 (2000).
[14] P. Dittrich and W. Banzhaf, in *Advances in Artificial Life*, Proceedings of the 6th European Conference on Artificial Life, Prague, 2001, edited by J. Kelemen and P. Sosik (Springer, Berlin, 2001), pp. 100–109.
[15] A. Soulier and T. Halpin-Healy, Phys. Rev. Lett. **90**, 258103 (2003).
[16] C. Kampmeier and B. Simon, J. Pers Soc. Psychol. **81**, 448 (2001).
[17] K. Taro Greenfeld, *Speed Tribes: Days and Nights with Japan's Next Generation* (Perennial, New York, 1995).
[18] J. M. McPherson and J. R. Ranger-Moore, Social Forces **70**, 19 (1991).
[19] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, Phys. Rev. E **68**, 046130 (2003).
[20] J. M. McPherson, L. Smith-Lovin, and J. Cook, Annu. Rev. Sociol. **27**, 415 (2001).
[21] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[22] D. J. Watts, Am. J. Sociol. **105**, 493 (1999).
[23] M. S. Granovetter, Am. J. Sociol. **78**, 1360 (1973).
[24] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).
[25] S. Janson, T. Łuczac, and A. Ruciński, *Random Graphs* (Wiley, New York, 1999).
[26] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).

[27] R. D. Bock and S. Z. Husain, Sociometry **13**, 146 (1950).
[28] L. Katz and J. H. Powell, Ann. Math. Stat. **28**, 442 (1957).
[29] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A **333**, 529 (2004).
[30] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[31] J. M. Roberts, Jr., Soc. Networks **22**, 273 (2000).
[32] F. Buckley and F. Harary, *Distance in Graphs* (Addison-Wesley, Redwood City, CA, 1989).
[33] B. Bollobás and F. R. K. Chung, SIAM J. Discrete Math. **1**, 328 (1988).
[34] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[35] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[36] A.-L. Barabási, R. Albert, H. Jeong, and G. Bianconi, Science **287**, 2115 (2000).
[37] W. Aiello, F. Chung, and L. Lu, *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (Association of Computing Machinery, New York, 2000), pp. 171–180.
[38] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103 (2002).
[39] F. Liljeros, C. R. Edling, L. A. Nunes Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).
[40] L. A. Nunes Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[41] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[42] P. Holme, C. R. Edling, and F. Liljeros, Soc. Networks **26**, 155 (2004).
[43] P. Bearman, J. Moody, and K. Stovel, Am. J. Sociol. **110**, 44 (2004).
[44] T. J. Fararo and M. H. Sunshine, *A Study of a Biased Friendship Net* (Syracuse University Press, Syracuse, NY, 1964).
[45] H. R. Bernard, P. D. Kilworth, M. J. Evans, C. McCarty, and G. A. Selley, Ethnology **27**, 155 (1988).
[46] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101 (2002).
[47] R. K. Leik and B. F. Meeker, *Mathematical Sociology* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[48] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[49] M. E. J. Newman, Phys. Rev. E **68**, 026121 (2003).
[50] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[51] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[52] M. E. J. Newman, Phys. Rev. E **67** 026126 (2003).
[53] M. E. J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).
[54] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[55] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).