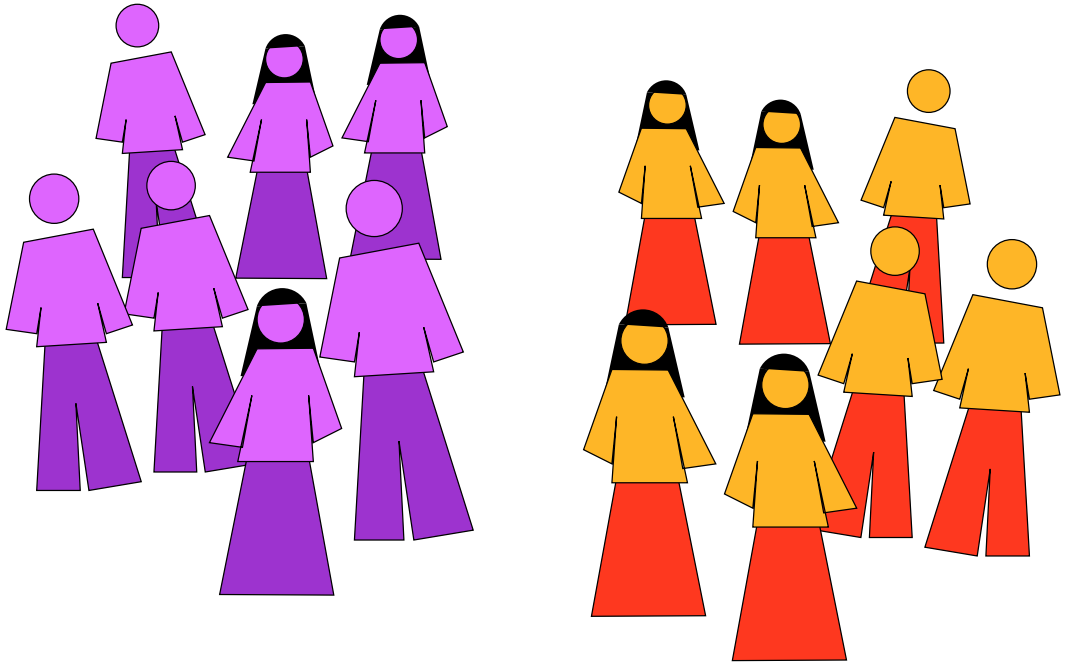
Modeling group formation in society: The networked seceder model

January 14, 2003

Petter Holme &
Andreas Grönlund
Department of Physics,
Umeå University, Sweden
NORDITA, Copenhagen,
Denmark

e-print cond-mat/0312010



LARGE SCALE SOCIAL NETWORK DATA

Affiliation networks

- Actor collaborations. $N = 449913$, $M = 25516482$ (Watts)
- Scientific collaborations. $N = 52909 - 1520251$, $M = 245300 - 11803064$ (Newman)
- Music collaborations. $N = 1275$ (Gleiser & Danon, de Lima e Silva *et al.*)
- Board of directors. $N = 7673$, $M = 55392$ (Davis *et al.*)

Online interaction

- E-mails. $N = 1700 - 59912$, $M = 15640 - 86300$ (Ebel *et al.*, Newman *et al.*, Guimerà *et al.*)
- Instant messaging. $N = 50259$, $M = 239452$ (Smith)
- Internet communities. $N = 29341$, $M = 115684$ (Holme *et al.*)
- Phone calls. $N = 47000000$, $M = 80000000$ (Aiello *et al.*)

Interviews

- Student romance. $N = 573$, $M = 477$ (Bearman *et al.*)
- Student friendship. $N = 417$ (Fararo *et al.*)
- Sexual contacts. $N = 2810$ (Liljeros *et al.*)
- Contact tracing. $N \sim 300$ (Potterat *et al.*)

Clustering coefficient

High in affiliation networks and interview based acquaintance networks. **Neutral** (= unbiased) in Internet communities, romantic and sexual networks.

Degree distribution

Power-law in telephone graphs, one e-mail study, sexual contacts. **Truncated power-law / stretched exponential** in Internet communities, Instant messaging, one e-mail studies and all affiliation data. **Exponential, Gaussian, Poissonian** Two e-mail studies, acquaintance networks based on interviews.

Degree-degree correlations

Positive in affiliation networks and acquaintance networks. **Neutral** in Internet communities, Instant messaging and one e-mail study.

Community (group) structure

Strong in affiliation networks, acquaintance networks, and (?) all other networks.

NETWORK MODELS OF GROUP FORMATION

Skyrms and Freemantle, PNAS **97**, 9340 (2000). Weighted network. The weight increase the more actors interact. Actors with higher weights are more likely to interact.

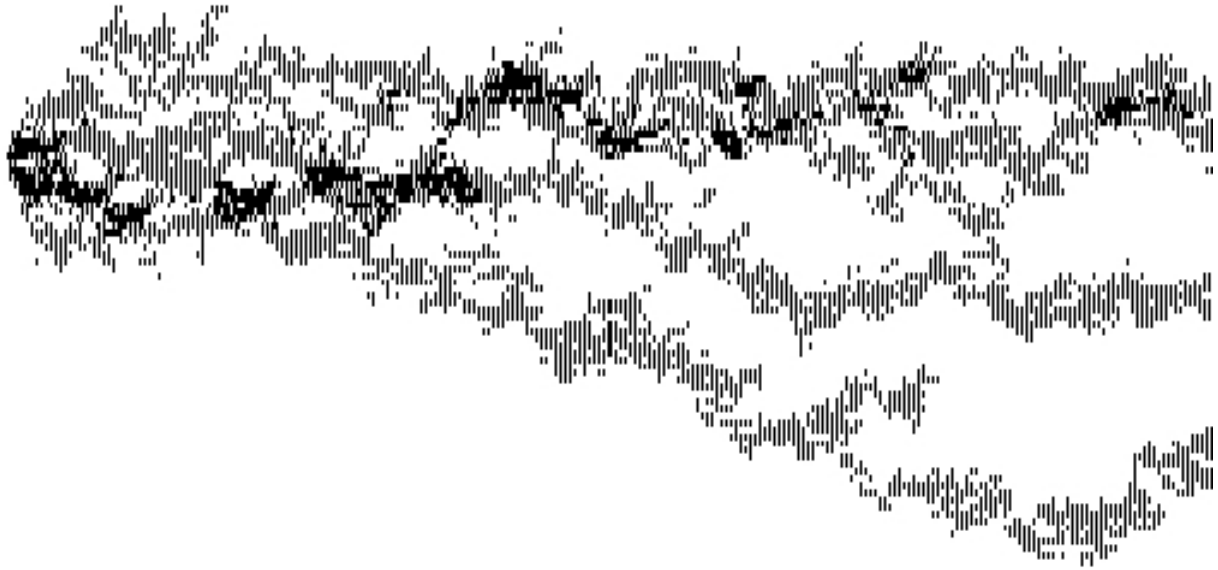
Jin, Girvan and Newman, PRE **64**, 046132 (2001). The degree cannot increase beyond a certain threshold. People with mutual friends are likely to get acquainted.

Motter, Nishikawa and Lai, PRE **68**, 036105 (2003). The traits of a person are given by a n-dimesional vector. Each dimension represent a characteristic social feature. Each dimension is hierarchically organized. Two vertices that are closer than a threshold value (in a certain metric) get an edge.

THE SECEDER MODEL

(Dittrich *et al.*, 2000.) N individuals with a real number $s(i)$ representing the traits 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 with $s(\hat{i}) + \eta$, where η is a random number from the normal distribution with mean zero and variance one.



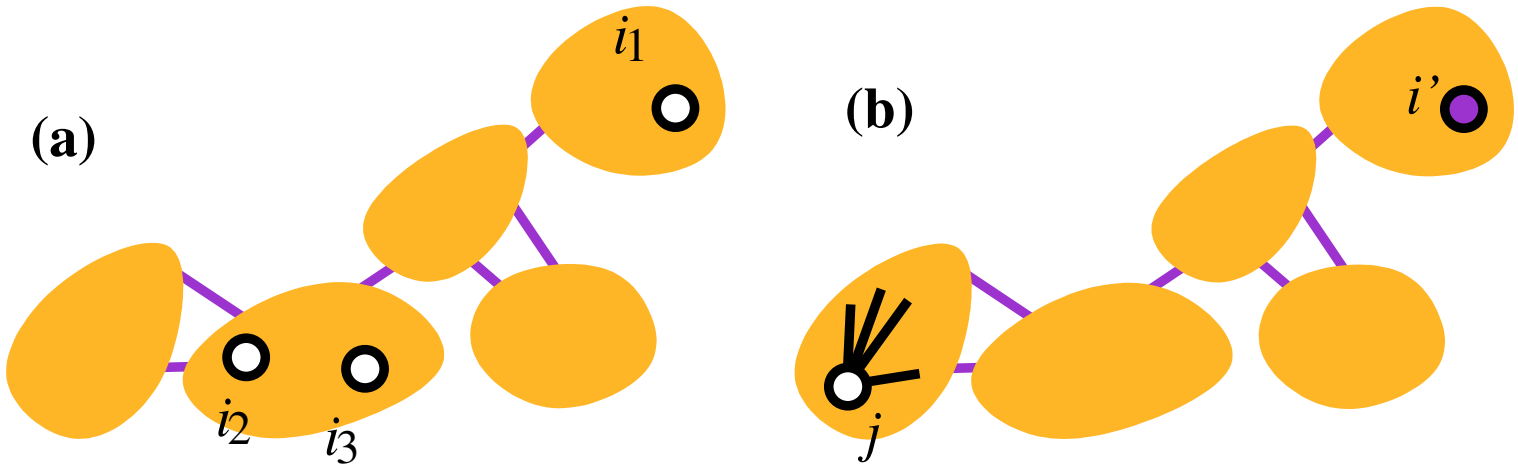
Generation

OUR NETWORKED SECEDER MODEL

Starting from any graph with N vertices and M edges we iterate the following steps:

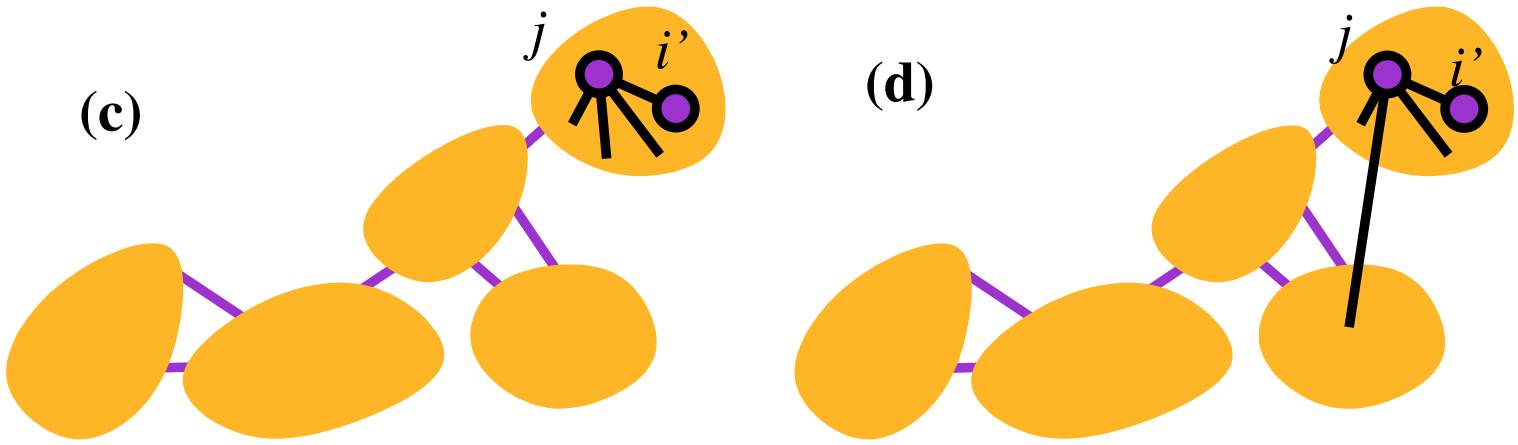
(a) Select three different vertices i_1, i_2 and i_3 with uniform randomness.

(b) 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 vertex in the set of least central vertices chosen uniformly randomly.

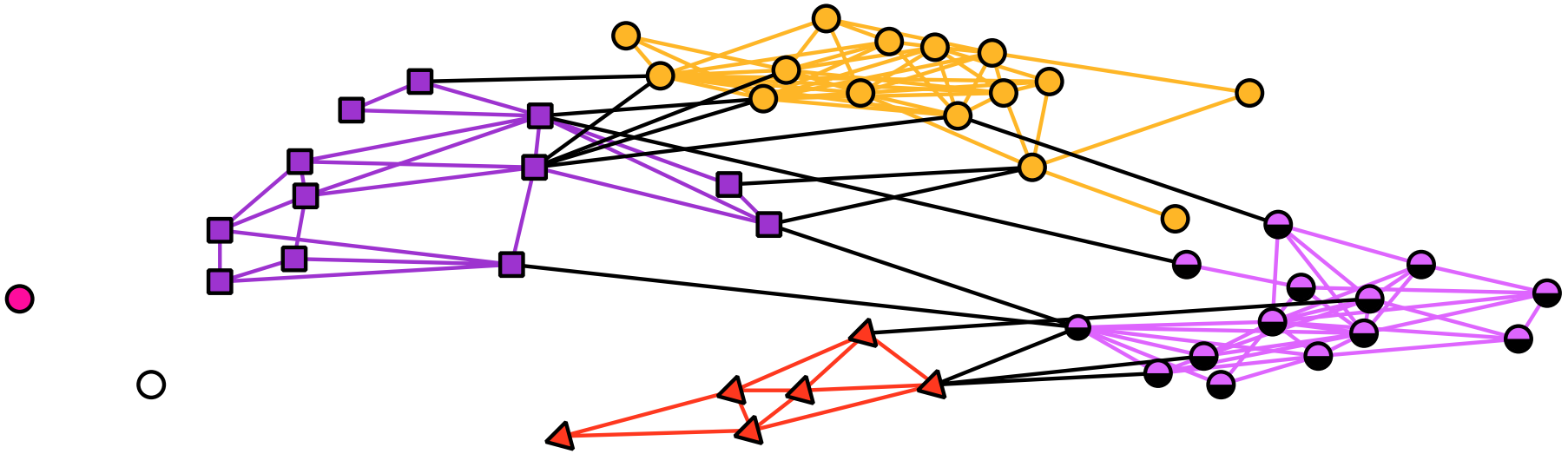


(c) Choose a vertex j by uniform randomness. If $\deg j \leqslant \deg \hat{i} + 1$, rewire j's edges to $\hat{i}$ and a random selection of $\hat{i}$'s neighbors. If $\deg j \geqslant \deg \hat{i} + 1$, rewire j's edges to $\hat{i}$, i's neighborhood and (if $\deg j > \deg \hat{i} + 1$) to $\deg j - \deg \hat{i} - 1$ randomly selected other vertices.

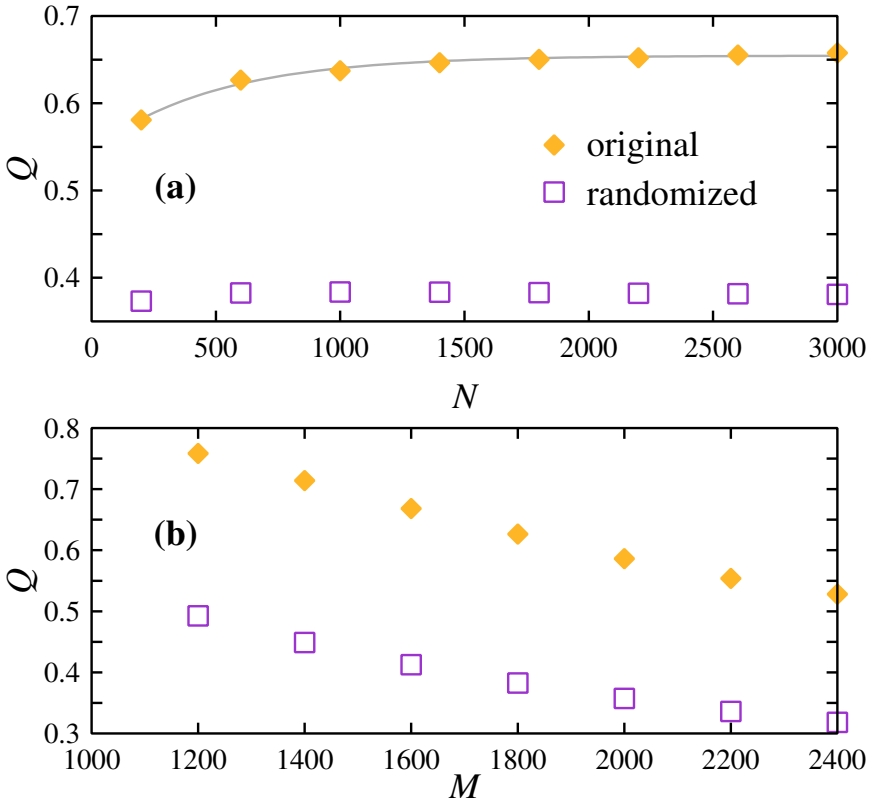
(d) Go through j's edges once more and rewire these with a probability p to a randomly chosen vertex.

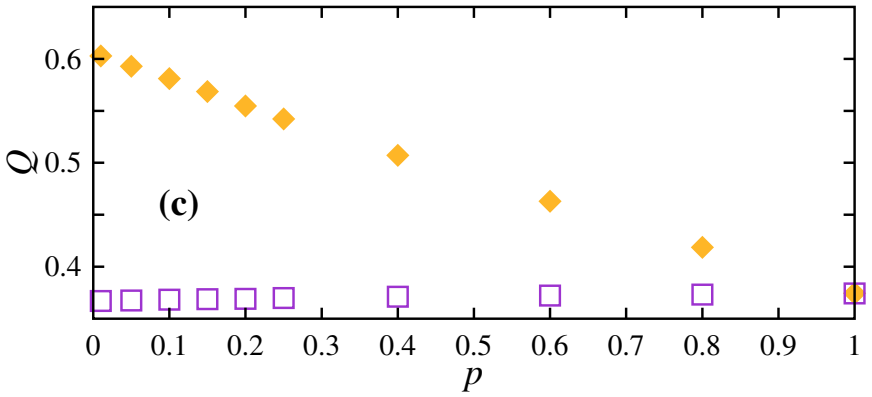


THE GROUP STRUCTURE



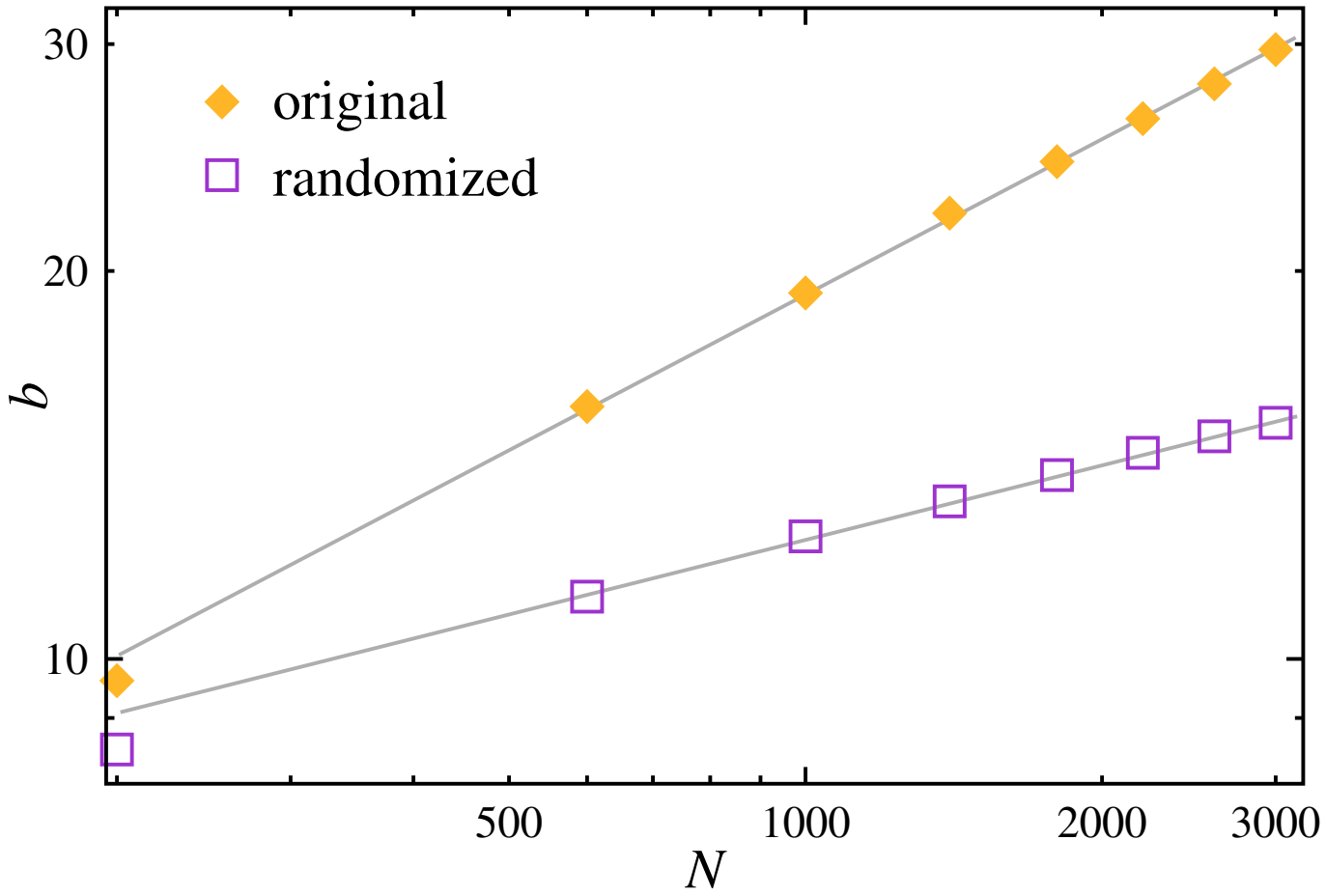
One realization of the networked seceder model. The model parameters are $N = 50$, $M = 150$ and $p = 0$. The indicated groups are identified with Newman's clustering algorithm (cond-mat/0309508). This realization have modularity $Q = 0.575$, clustering coefficient $C = 0.530$, and assortative mixing coefficient $r = 0.0456$.



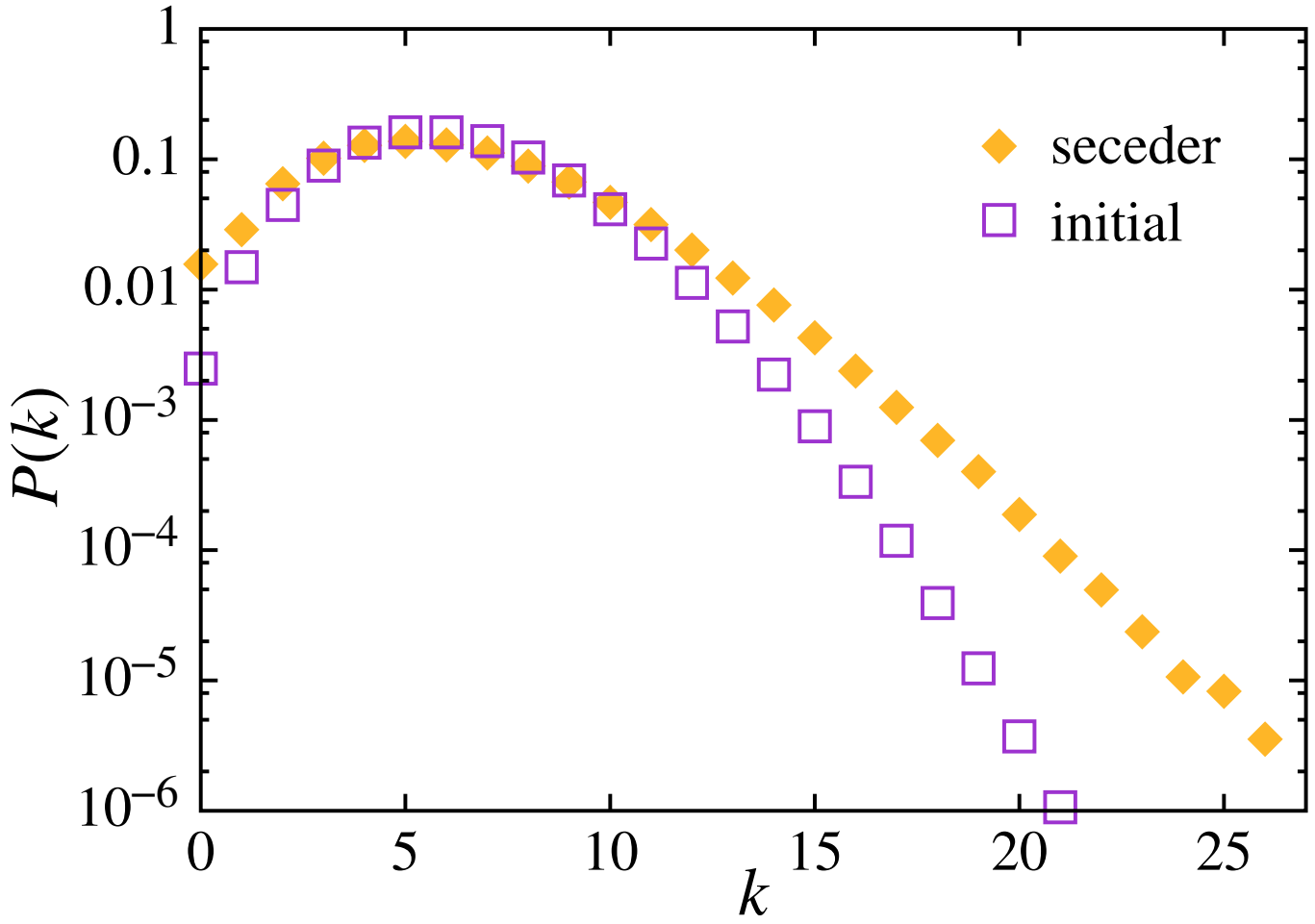


$$Q' = \sum_{s \in S} (e_{ss} - a_s^2)$$

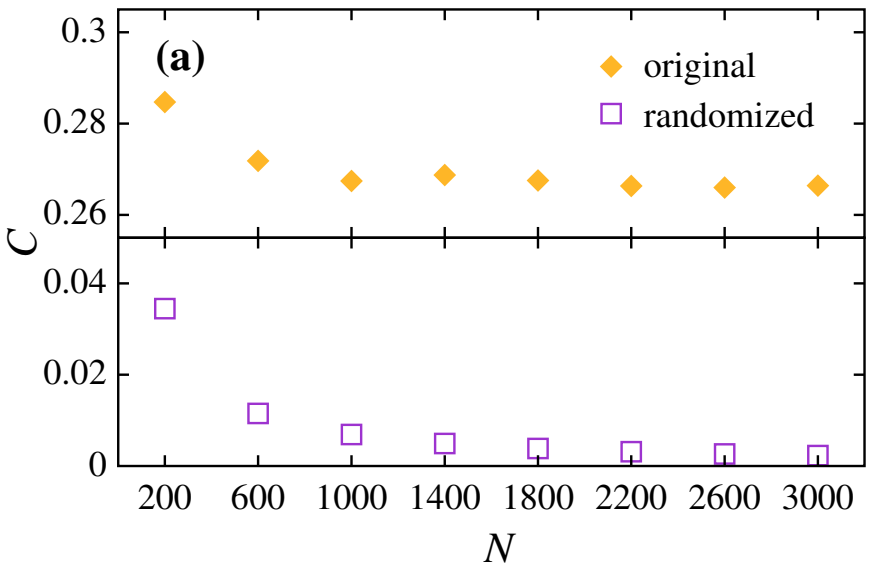
where S is the set of subnetworks at a specific iteration of the algorithm and $e_{ss'}$ is the fraction of edges that goes between a vertex in s and a vertex in s', and $a_s = \sum_{s'} e_{ss'}$.

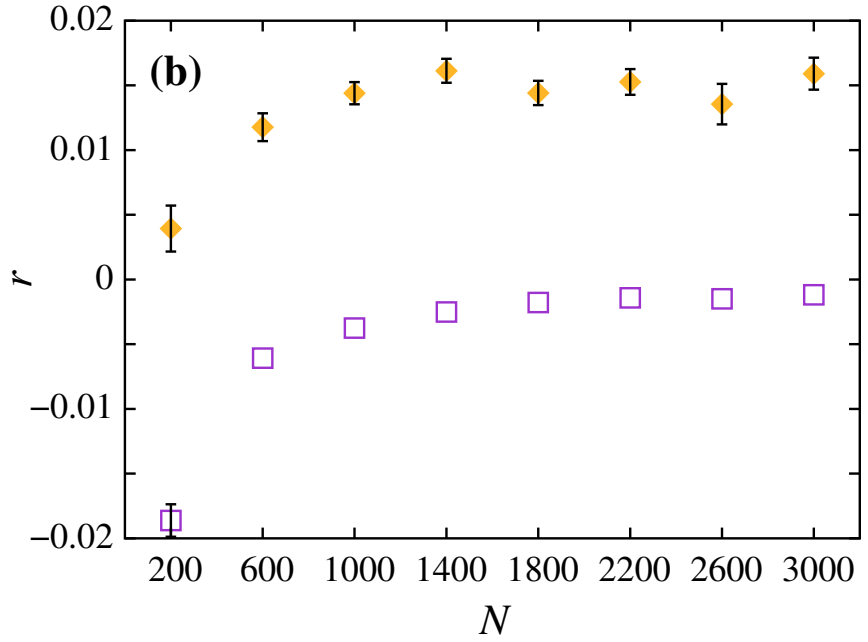


The number of groups b as a function of the systems size N. The other parameter values are $M = 3N$ and $p = 0.1$. The line is a fit to a power-law ab^{β}.



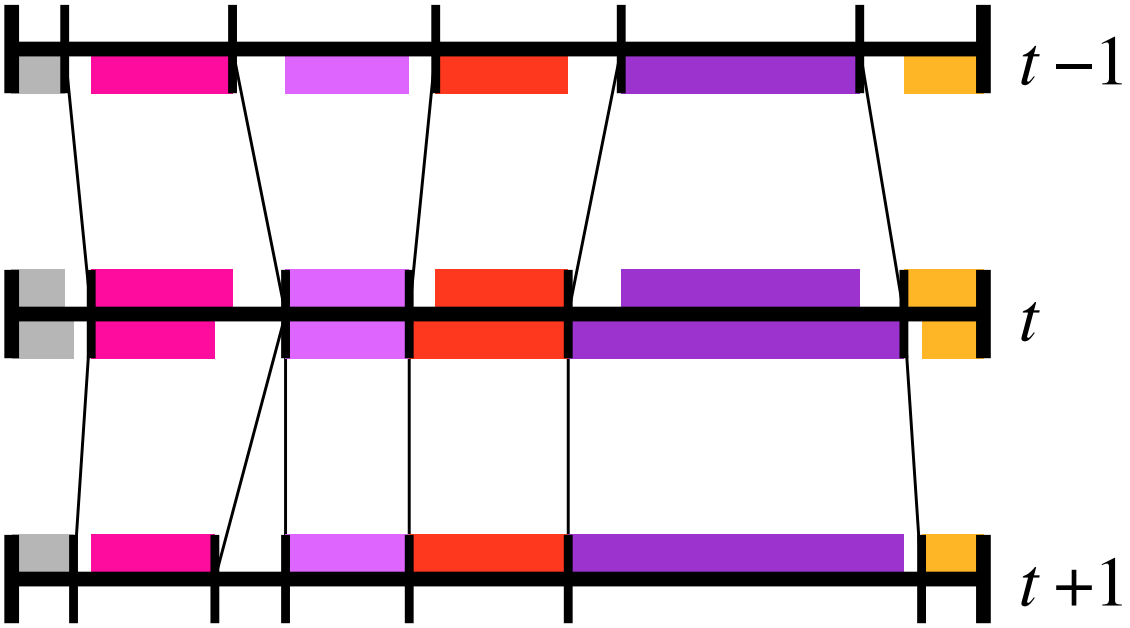
Degree distribution of the networked seceder model. The model parameters are $N = 1800$, $M = 5400$ and $p = 0.1$. The squares indicate the degree distribution of a random graph with the sizes (N and M), i.e., the initial network before the iterations of the seceder model commence.



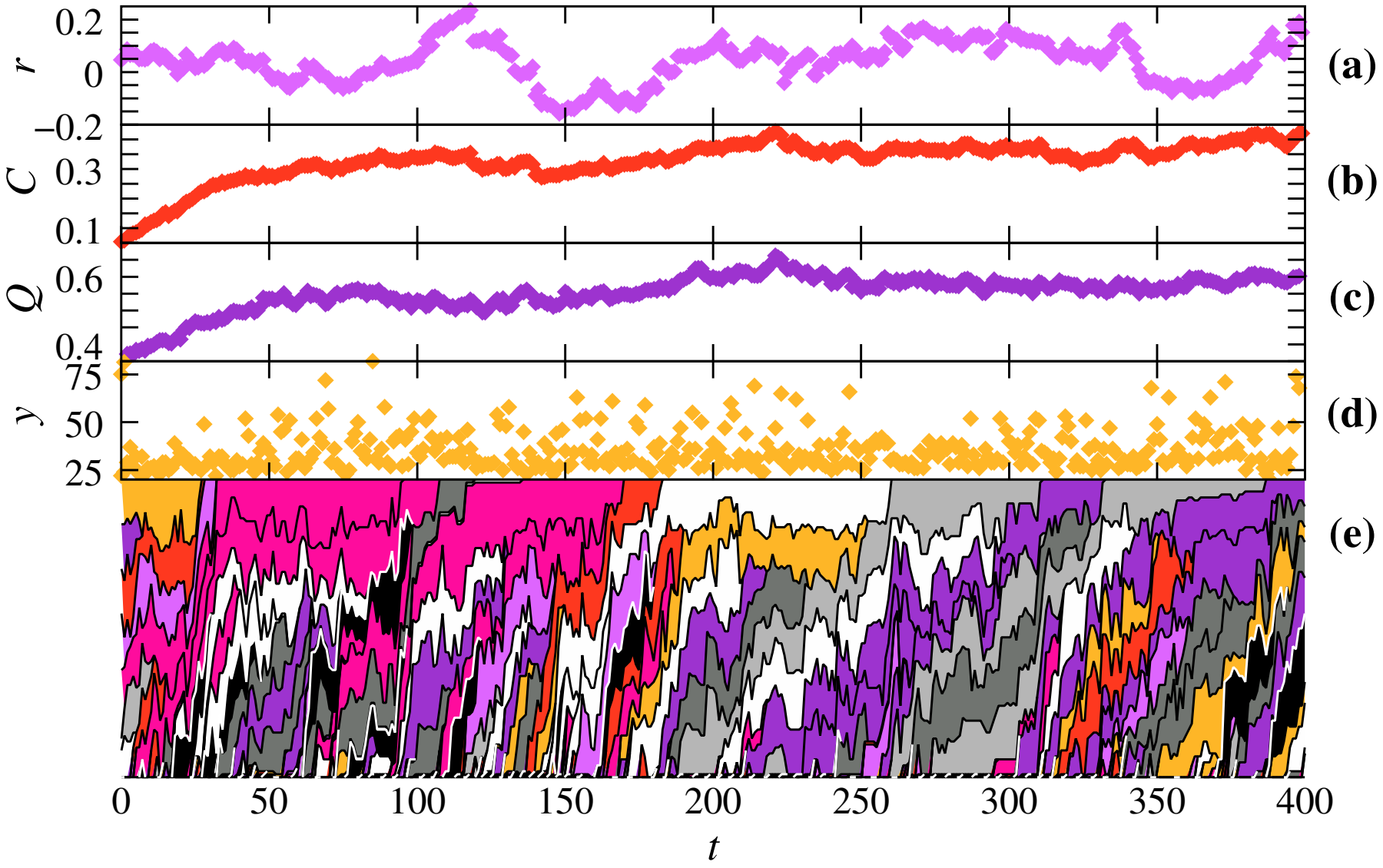


(a) Clustering coefficient is finite. **(b)** Degree-degree correlations are positive.

TIME EVOLUTION OF THE COMMUNITIES



Communities at consecutive time steps are identified by maximizing the overlap.


r
0.2
0
−0.2
C
0.3
0.1
Q
0.6
0.4
y
75
50
25
(a)
(b)
(c)
(d)
(e)
0
50
100
150
200
250
300
350
400
t