Reply to "Comment on 'Subgraphs in random networks' "

S. Itzkovitz,[1] R. Milo,[1] N. Kashtan,[1] M. E. J. Newman,[2] and U. Alon[1]
[1]*Department of Molecular Cell Biology and Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 7600, Israel*
[2]*Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109, USA*
(Received 20 May 2004; published 18 November 2004)

King [preceding Comment, Phys. Rev. E **70**, 058101 (2004)] points out biases in one of the two common algorithms for generating simple random graphs—the matching, or stub-pairing, algorithm. We clarify that in our simulations of simple graphs we used a different algorithm, the Markov-chain Monte Carlo switching algorithm, which is more uniform. As for multigraphs, the stub-pairing algorithm indeed samples uniformly configurations rather than multigraphs, as King points out, and thus is relevant for our model, which pertains to configurations. Finally, we demonstrate that the algorithm we used to generate families of random networks with scale-free out-degree and compact in-degree does not result in noticeable biases.

DOI: 10.1103/PhysRevE.70.058102 PACS number(s): 89.75.Hc

King [1] correctly points out that the direct network generating algorithm used by Newman *et al.* and many others (matching, or stub-pairing, algorithm) samples graphs with multiple edges between vertex pairs with lower probability than simple graphs. This bias did not affect the calculations of Newman *et al.* [2], since they were interested in graph properties only in the large system size limit, and the corrections due to multiple edges vanish in this limit.

We presented approximate formulas for subgraph numbers on networks drawn from the configuration model—multigraphs with an arbitrary degree sequence [3]. We showed that for many real-world networks these formulas give a good estimate also for random graphs with an arbitrary degree sequence, with only single edges allowed.

King points out that the modified matching algorithm presented in [4] does not uniformly sample simple graphs (simple graphs are graphs that are constrained not to have multiple edges in the same direction between two nodes). As a result of King's Comment, we have recently tested the degree of nonuniformity of various randomizing algorithms [5]. Two methods, a Markov-chain Monte Carlo (switching) algorithm and the modified matching algorithm, were benchmarked against a different and more uniform non-Markov-chain algorithm ("go with the winners" [6,7]). It was shown on both a toy model network and on networks with the degree sequence of real-world networks that the switching method samples graphs with single edges as uniformly as the "go with the winners" algorithm, and does so with a relatively fast mixing time. The modified matching algorithm introduces a bias which becomes more prominent on networks in which both out-degree and in-degree sequences are heavy tailed. We therefore recommend the switching algorithm for purposes of network motif detection and network analysis. All the results in [4,8], as well as the direct enumeration results for graphs with single edges of Table I in [3], were obtained using the switching algorithm.

To test our approximate formulas for the *configuration model* (which allows multiple edges), we used the matching algorithm in Table I of [3], allowing more than one edge between two nodes. This resulted in good agreement with our approximate formulas for multiple-edge networks (Table I of [3]). King correctly points out that the matching algorithm uniformly generates configurations and not multigraphs. Indeed, our approximate formulas relate to configurations and not to multigraphs. The formulas rely on the fact that the expected number of edges between a node with K outgoing edges and a node with R incoming edges is $K*R/E$, where E is the number of edges in the network. In the toy network of Fig. 1 of King, this predicts $2*2/4=1$ edge between nodes W and Z. Indeed, averaging the number of edges between nodes W and Z over the 24 *configurations* results in the expected average of 1. Averaging the number of edges between nodes W and Z over the seven *multigraphs*,

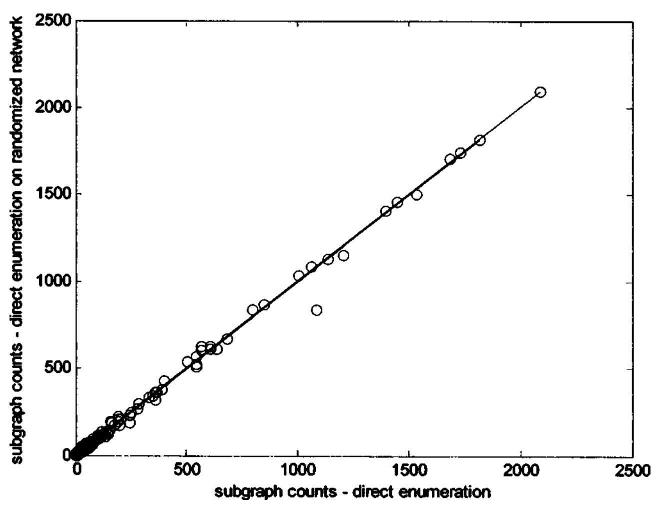

FIG. 1. Comparison of subgraph counts generated by two processes: (1) the algorithm described in the text for creating an ensemble of scale-free random networks and (2) a switching algorithm applied to each network of (1), with a further randomization of 100–200 switches per edge. Shown are the numbers of feed-forward loop subgraphs in 1000 random networks with $N=2000$ nodes, with scale-free out-degree ($\gamma=2.1$) and compact in-degree.

however, results in a higher number (8/7) (each one of the top two multigraphs in King's Fig. 1 includes two configurations, and each one of the bottom five graphs includes four).

In Ref. [3], we further tested the scaling of subgraph numbers, based on the approximate formulas, on an ensemble of random networks with a scale-free out-degree and compact in-degree (Figs. 3 and 5 of Ref. [3]). To construct these networks from scratch, we used the following algorithm: For each node i a number W_i was drawn from a scale-free distribution, and then each node of the remaining $N-1$ network nodes was connected to node i with a single edge if: $n_r < W_i$, where n_r is a random number uniformly generated between 0 and 1. This results in an out-degree of W_i on average for node i, and in a compact in-degree for all nodes. In this class of random networks, the probability of more than a single edge between any two nodes is vanishingly small in the large system size limit and the uniformity bias is negligible. Further randomizing these networks using the switching algorithm did not change the results (Fig. 1).

We thank Oliver King for motivating the research presented in [5].

[1] O. D. King, preceding Comment, Phys. Rev. E **70**, 058101 (2004).
[2] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. E **64**, 026118 (2001).
[3] S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, and U. Alon, Phys. Rev. E **68**, 026127 (2003).
[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U Alon, Science **298**, 824 (2002).
[5] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, e-print cond-mat/0312028.
[6] D. Aldous, and U. V. Vazirani, IEEE Symposium on Foundations of Computer Science (1994), pp. 492–501 (unpublished).
[7] P. Grassberger and W. Nadler, e-print cond-mat/0010265.
[8] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M Sheffer, and U. Alon, Science **303**, 1538 (2004).