Supplementary Information

for

Existence of outsiders as a characteristic of online communication networks

Takeguchi, Maehara, Toyoda, and Kawarabayashi

published in *Network Science*

S1 Rewiring of links connected to outsiders

We first remove all the links connected to outsiders and then rewire the links while keeping the degree of all the nodes. In total, $46,465$ links associated with outsiders are rewired ($\sim 5\%$ of $M = 927,352$ links in the original mention network). This rewiring process keeps the total number of links and the degree distribution (it possibly create self-loops but this effect is negligible since the number of the rewired links is sufficiently large). By contrast, the rewiring process may change C_i, r, and the connectedness of the network. If it holds true that outsiders cast their links to other nodes in a random manner, the following properties should be observed for the rewired network: (1) an increase in r after the removal of the same set of outsiders, (2) low connectivity between outsiders, (3) similar V_i and C_i values for outsiders as those in the original Mention network.

In Fig. S1, the results for a rewired network are shown, which indicate that the links between outsiders and other nodes are not completely random. First, as shown in Fig. S1(a), the r value of the rewired network without node removal is equal to 0.341 and larger than 0.135 for the original Mention network. This implies that outsiders help the rewired network be assortative, which is also supported by the fact that r value decreases in $f_{\text{removed}} \lesssim 0.002$. The s_{removed} value in the rewired network is larger than the original network, and outsiders are more connected to each other. The rewiring process reduces the variety of V_i and C_i among outsiders as shown in Figs. S1(b) and (c). While V_i of outsiders take values within $[0.3, 2.5]$ in the original network, V_i in the rewired network concentrate around unity. In a similar way, V_i of outsiders take values within $[0, 0.3]$, C_i in the rewired network take value close to zero, regardless the V_i values in the original network. These results suggest that the links between outsiders and other nodes play an important role in determining the local structure around outsiders and that the links are not created in a random manner.

S2 Ratio between the retweets and tweets

In Fig. 5 of the main text, we plotted the total number of tweets and times being retweeted, denoted by n_i^{T} and n_i^{R} for node i, respectively. Figure S2 shows the ratio of the two quantities, i.e., $n_i^{\text{R}}/n_i^{\text{T}}$, which gives the expecting number of times being retweeted per tweet, for outsiders and non-outsiders. This plot implies that outsiders are similar to the rest of users in terms of this quantity.

S3 Results for other data sets

We examine the outsider model in other network data sets to investigate the generality of the existence of outsiders that we confirmed in the Mention network. The five network data sets we use are the Enron [Klimt & Yang, 2004], EU-email [Leskovec *et al.*, 2007], Facebook [Viswanath *et al.*, 2009], Slashdot [Gómez *et al.*, 2008], and Wiki-talk [Leskovec *et al.*, 2010] networks that were all originally made available online by the Koblenz Network Collection (http://konect.uni-koblenz.de/)[Kunegis, 2013]. In all of these networks, we define the nodes by the users of the online communication tools and the links by the reciprocal interactions between pairs of users. The Enron and EU-email are based on the record of email exchanges within a company and an European institution. The Facebook network is based on the the post of messages by a user on another user's personal page in Facebook. The Slashdot and Wiki-talk networks are based on the message exchanges between the users of these websites. The basic statistics of the five networks are summarized in Tab. S1 The results for the five networks are shown in Figs. S3, S4, S5, S6, and S7. We fix the number of outsiders $N_{\text{outsider}} = 1,000$ for all the networks.

For the Facebook network, in which we observe the outsider properties (Fig. S5) similar to those found in the mention network, the popularity of outsiders and the rest of users are compared. Figure S8 shows the histograms of the total number of posts and subscribed friends for outsiders and non-outsiders. It should be noted that the total number of subscribed friends of a user is in general different from the node degree in the Facebook network, because the Facebook network is defined by the message exchange between users, not the friendship subscription.

References

[Gómez *et al.*, 2008] Gómez, V., Kaltenbrunner, A., & López, V. (2008). Statistical analysis of the social network and discussion threads in slashdot. *Pages 645–654 of: Proceedings of the 17th International Conference on World Wide Web*. Beijing, China: ACM.

[Klimt & Yang, 2004] Klimt, B., & Yang, Y. (2004). The Enron corpus: A new dataset for Email classification research. *Pages 217–226 of: Machine Learning: ECML 2004 Lecture Notes in Computer Science Volume 3201*. Heidelberg, Germany: Springer Berlin Heidelberg.

[Kunegis, 2013] Kunegis, J. KONECT: the Koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web Companion.*, pp. 1343–1350, Rio de Janeiro, Brazil, ACM (2013).

[Leskovec *et al.*, 2007] Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Desification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, **1**(1), 2.

[Leskovec *et al.*, 2010] Leskovec, J., Huttenlocher, D. P., & Kleinberg, J. M. (2010). Governance in social media: A case study of the Wikipedia promotion process. *Pages 98–105 of: Proceedings of the Forth International Conference on Weblogs and Social Media*. Washington, D.C.: AAAl.

[Viswanath *et al.*, 2009] Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). On the evolution of user interaction in Facebook. *Page 37 of: Proceedings of the Second ACM Workshop on Online Social Networks*. Barcelona, Spain: ACM.

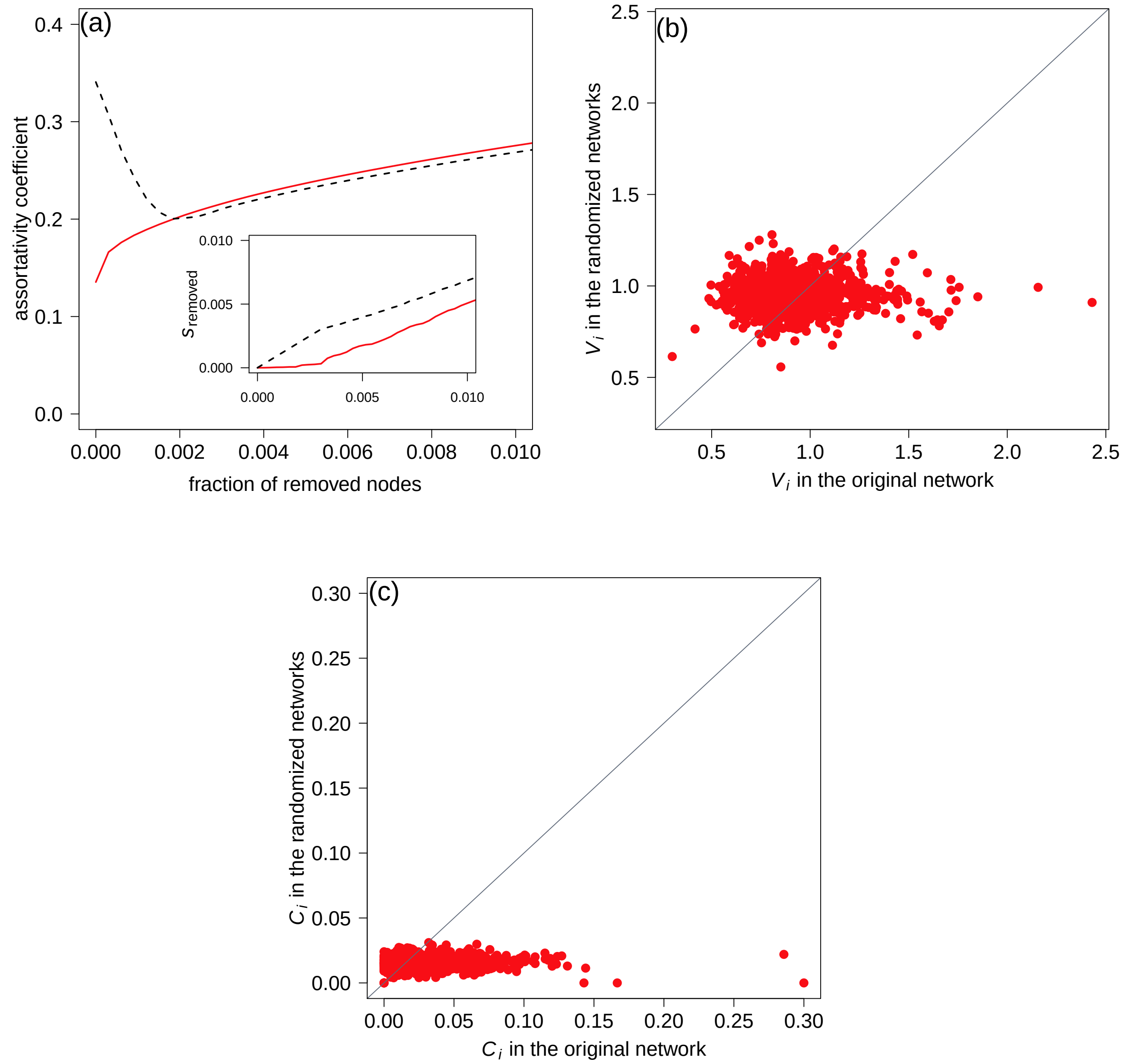

Figure S1: Results for a rewired mention network. (a) Degree assortativity coefficient r of the resultant networks (main panel) and the sizes of the largest connected component of the removed nodes s_{removed} (inset), as a function of the proportion of nodes removed f_{removed}. The dashed lines are the results for the rewired network and the solid lines the original network as a reference. (b) Scatter plot of the coefficient of variation of the neighbors' degree V_i of outsiders in the original network (horizontal axis) against V_i in the rewired network (vertical axis). Each dot corresponds to an outsider. (c) Scatter plot of the local clustering coefficient C_i of outsiders in the original network (horizontal axis) against C_i in the rewired network (vertical axis).

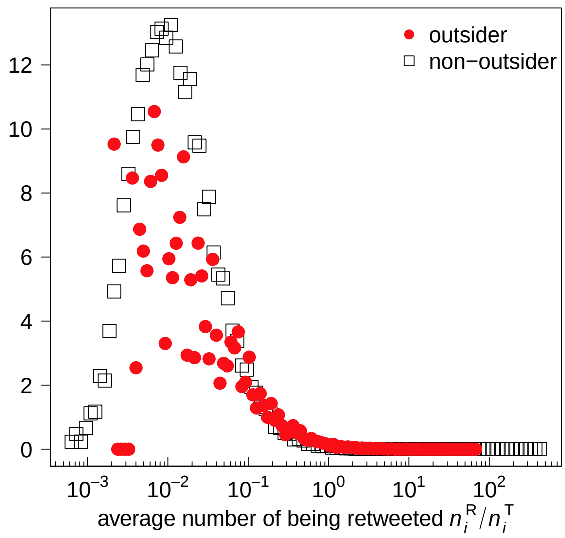

Figure S2: Histograms of the ratio of the total number of times being retweeted to the total number of tweets, $n_i^{\mathrm{R}}/n_i^{\mathrm{T}}$, for outsiders (bullets) and non-outsiders (squares).

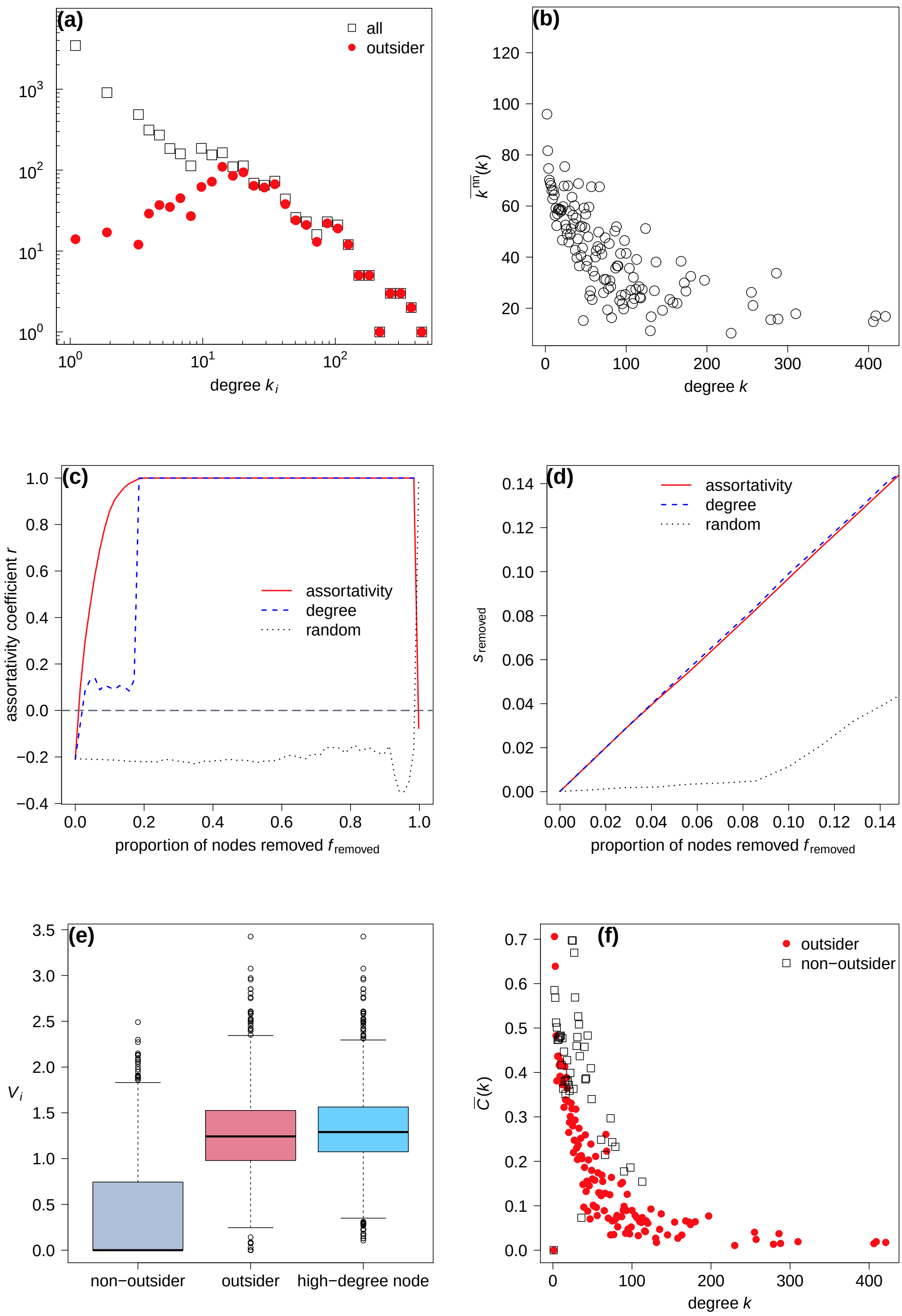

Figure S3: Results for the Enron network. (a) Histogram showing the node degree for outsiders and non-outsiders. (b) The average degree of nodes adjacent to the nodes with degree k. (c) Assortativity coefficient r (solid lines) and (d) the sizes of the largest connected component of removed nodes s_{removed}, as a function of the proportion of nodes removed f_{removed}. (e) Diversity of neighbors' degree V_i. (f) Average local clustering coefficient $\overline{C}(k)$ as a function of node degree k.

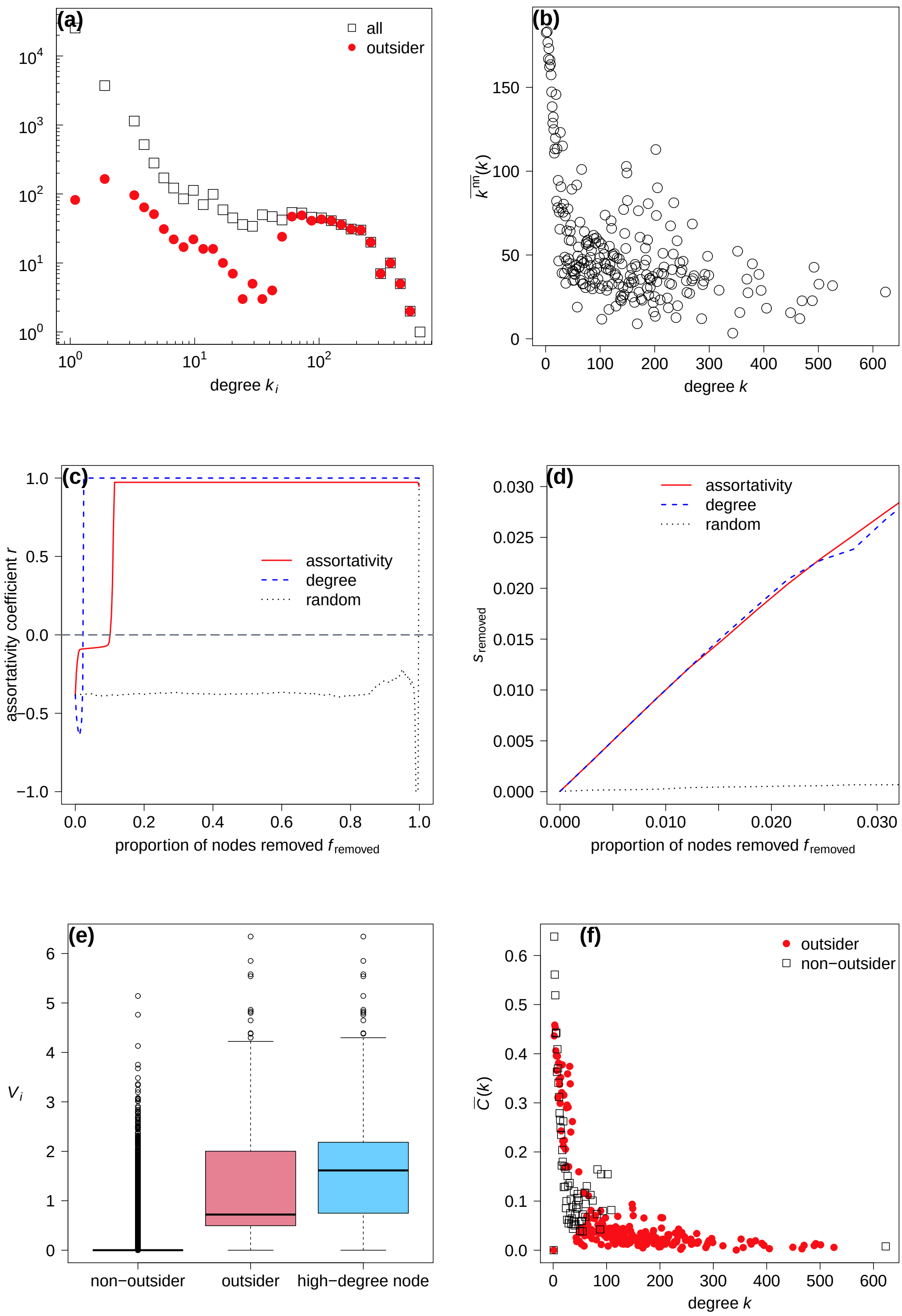

Figure S4: Results for the EU-email network. (a) Histogram showing the node degree for outsiders and non-outsiders. (b) The average degree of nodes adjacent to the nodes with degree k. (c) Assortativity coefficient r (solid lines) and (d) the sizes of the largest connected component of removed nodes s_{removed}, as a function of the proportion of nodes removed f_{removed}. (e) Diversity of neighbors' degree V_i. (f) Average local clustering coefficient $\overline{C}(k)$ as a function of node degree k.

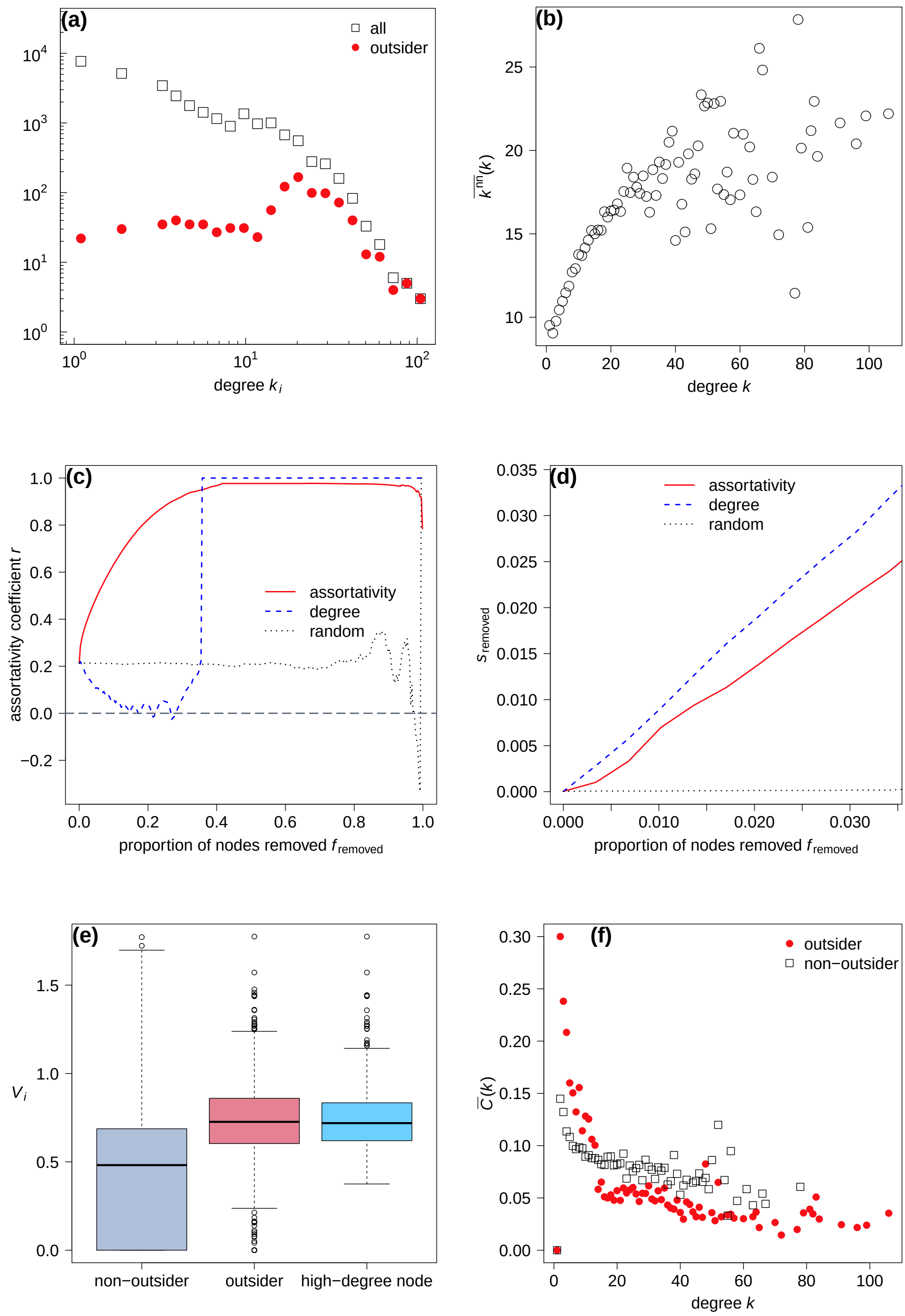

Figure S5: Results for the Facebook network. (a) Histogram showing the node degree for outsiders and non-outsiders. (b) The average degree of nodes adjacent to the nodes with degree k. (c) Assortativity coefficient r (solid lines) and (d) the sizes of the largest connected component of removed nodes s_{removed}, as a function of the proportion of nodes removed f_{removed}. (e) Diversity of neighbors' degree V_i. (f) Average local clustering coefficient $\overline{C}(k)$ as a function of node degree k.

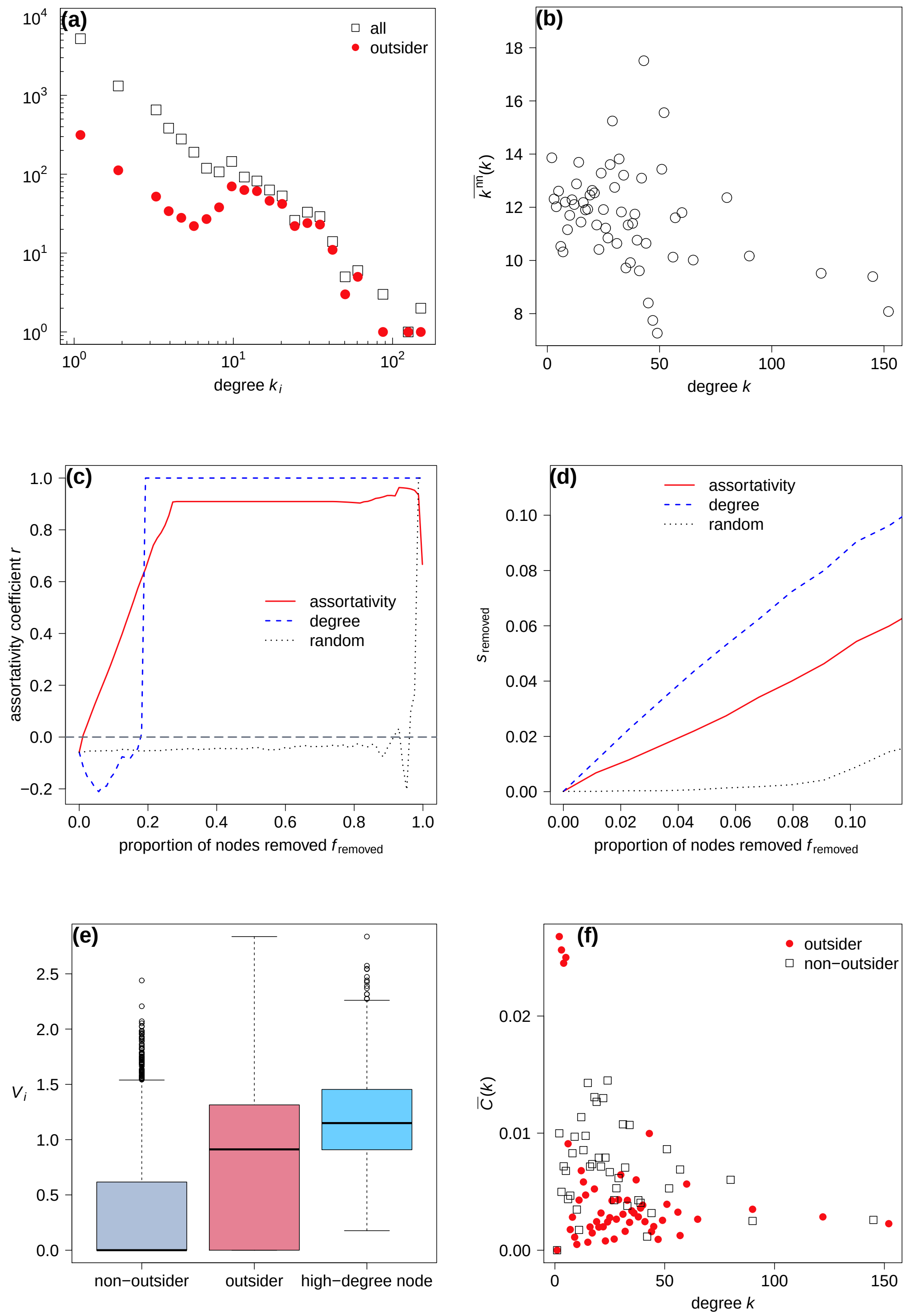

Figure S6: Results for the Slashdot network. (a) Histogram showing the node degree for outsiders and non-outsiders. (b) The average degree of nodes adjacent to the nodes with degree k. (c) Assortativity coefficient r (solid lines) and (d) the sizes of the largest connected component of removed nodes s_{removed}, as a function of the proportion of nodes removed f_{removed}. (e) Diversity of neighbors' degree V_i. (f) Average local clustering coefficient $\overline{C}(k)$ as a function of node degree k.

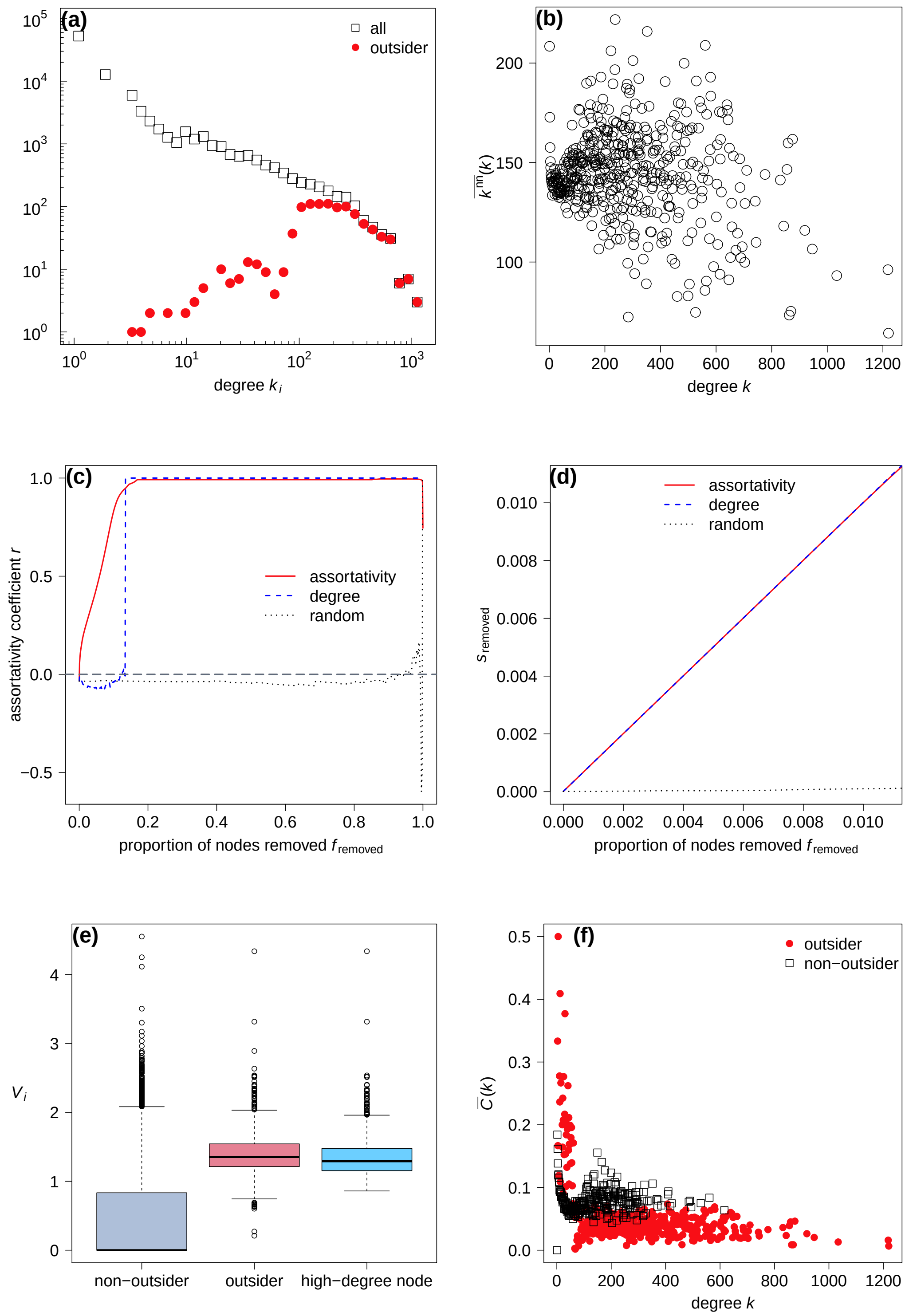

Figure S7: Results for the Wiki-talk network. (a) Histogram showing the node degree for outsiders and non-outsiders. (b) The average degree of nodes adjacent to the nodes with degree k. (c) Assortativity coefficient r (solid lines) and (d) the sizes of the largest connected component of removed nodes s_{removed}, as a function of the proportion of nodes removed f_{removed}. (e) Diversity of neighbors' degree V_i. (f) Average local clustering coefficient $\overline{C}(k)$ as a function of node degree k.

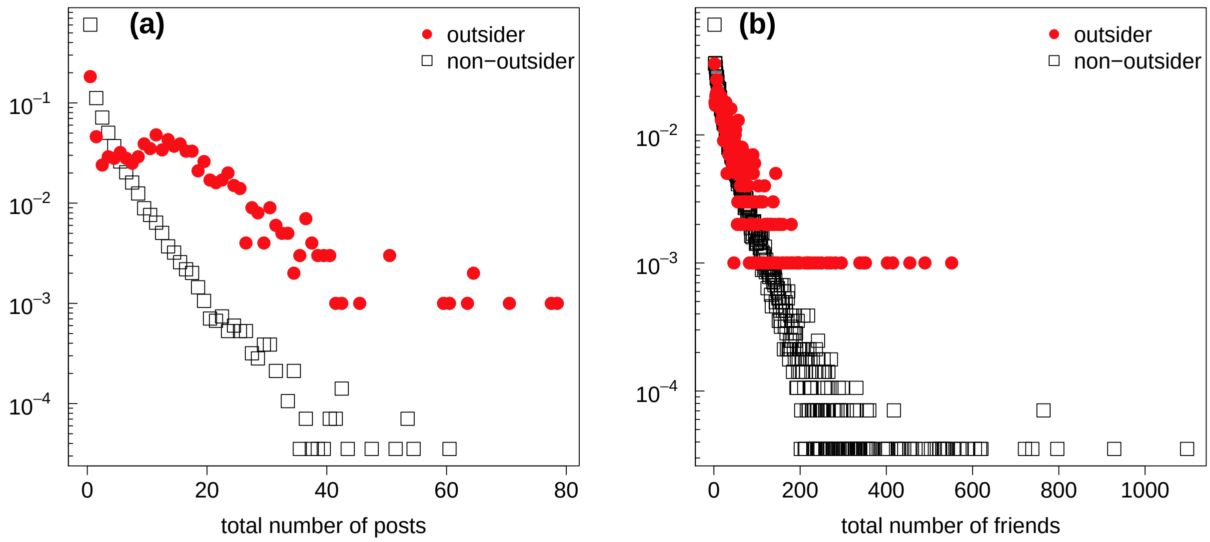

Figure S8: Histograms of the total number of (a) posts and (b) subscribed friends for outsiders (bullets) and non-outsiders (squares) of the Facebook network.

Table S1: Summary of basic statistics of the network data sets used: the total number of nodes N and links M, the average clustering coefficient C, and the degree assortativity coefficient r.

Name	N	M	C	r
Enron	7,015	22,474	0.240	−0.209
EU-email	32,430	54,397	0.113	−0.382
Facebook	29,342	79,230	0.084	0.213
Slashdot	8,815	12,859	0.003	−0.058
Wiki-talk	92,117	360,767	0.059	−0.034