Necessary backbone of superhighways for transport on geographical complex networks

Yukio Hayashi[1]

Japan Advanced Institute of Science and Technology, 1-1 Asahidai Ishikawa, 923-1292 Japan, yhayashi@jaist.ac.jp

Summary. We show drastic improvement of the robustness in geographical networks by adding a small fraction of shortcuts. We investigate whether shortcuts highly contribute to transfer packets according to the large centrality of links. Spatially heterogeneous communication requests on naturally emerged dense and sparse areas are also discussed.

1 Introduction

Real complex networks, such as power-grid, airline flight-connection, and the Internet, are **embedded in a metric space**, and **long-range links are restricted** [1, 2] for economical reasons. Moreover, there exist common topological characteristics: one is the small-world (SW) phenomenon that each pair of nodes are connected through relatively small number of hops to a huge network size N, another is the scale-free (SF) structure that follows a power-law degree distribution $P(k) \sim k^{-\gamma}$, $2 < \gamma < 3$, which consists of many nodes with low degrees and a few hubs with high degrees. Intuitively, a path through hubs are short because of many possibilities to select proper mediators or a terminal node, thus the SW property is also maintained. Although such geographical SF network models have been gradually considered, unfortunately, the error tolerance of connectivity becomes **more vulnerable in a spatial construction** than that in the assumption of tree-like structure. Indeed, in a theoretical prediction [3, 4], the percolation threshold is increased by the majority of small-order cycles that locally connected with a few hops. As the smallest-order, triangular cycles tend to be particularly constructed by a geographical constraintt to neighbor nodes.

In contrast, it has been suggested that higher-order cycles connected with many hops improve the robustness in the theoretical analysis on a one-dimensional SW model modified by adding shortcuts between two nodes out of the connected neighbors [5]. Recently, it has been numerically shown [6] that the robustness is improved by fully random rewirings under the same degree distributions in typical geographical network models: Delaunay triangulation (DT) [7], random Apollonian (RA), and Delaunay-like scale-free (DLSF) networks [6, 8]. As similar to connecting local areas by rewirings without geographical constraints, we expect **the shortcut effect on the improvement of robustness** in such geographical SF networks. Adding shortcuts is practically more natural rather than rewirings, because the already constructed links are not wastefully discarded.

Essentially, between isolated clusters, any packets are unable to be transferred. Thus, it is important to find such **a necessary backbone for maintaining the whole connectivity**. However, the optimization is generally difficult, since the minimum dominating set problem is NP-hard [9]. This criterion aims to reduce the number of base-stations with high load on which many packets are concentrated. Instead of the minimum, we consider a candidate of the necessary backbone for transport in a viewpoint of network science, e.g. by using percolation analysis and the centrality of links.

2 Geographical network models

Planar networks without crossing of links are suitable for efficient geographical routings on wired and/or wireless connections, since we can easily find the shortest distant path from a set of edges on the faces that intersect the straight line between the source and terminal. In computer science, online routing algorithms [10] that **guarantee delivery of messages using only local information** about positions of the source, terminal, and the adjacent nodes to a current node are well-known. As a connection to SF networks, we consider Delaunay triangulation (DT) and random Apollonian (RA) network models based on planar triangulation of a polygonal region. DT is the optimal planar triangulation in some geometric criteria [7], and the ratio of the shortest path length is bounded by a constant factor to the direct Euclidean distance between any source and terminal [11], while RA network belongs to both SF and planar networks [12, 13], however long-range links inevitably appear near the edges of an initial polygon. To reduce the long-range links, Delaunay-like scale-free (DLSF) network has been proposed [6, 8]. Figure 1 shows the typical structure of each network, whose spatial arrangement emerges with **mixing of dense and sparse areas as similar to a population density**.

These planar networks are constructed as follows.

Step 0: Set an initial planar triangulation on a space.

Step 1: At each time step, select a triangle at random and add a new node at the barycenter. For each model, different linking processes are applied.

- RA: Then, connect the new node to its three nodes as the subdivision of triangle.
- DLSF: Moreover, by iteratively applying diagonal flips [7], connect it to the nearest node within a radius defined by the distance between the

new node and the nearest node of the chosen triangle, as shown in Figure 2.
If there is no nearest node within the radius, this flipping is skipped, therefore the new node is connected to the three nodes.

DT: After the subdivision of the chosen triangle, diagonal flips are globally applied to any pair of triangles until the minimum angle is not increased by exchanging diagonal links in the quadrilateral.

Step 2: The above process is repeated until the required size N is reached.

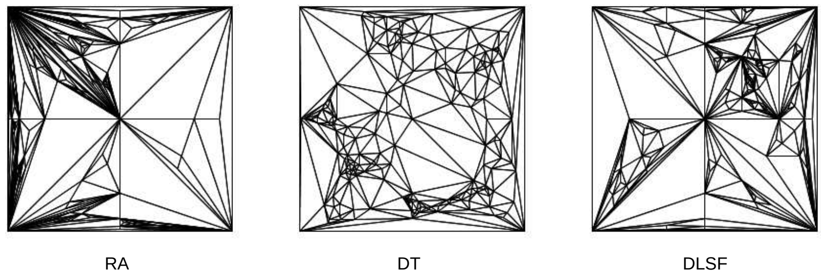


Fig. 1. Example of geographical networks grown from an initial triangulation of square to a configuration with mixing of dense and sparse areas as similar to a population density. Note that the four corners and the center points are hubs in RA and DLSF networks.

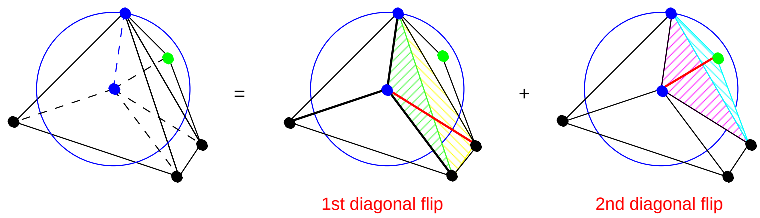


Fig. 2. Linking procedures in a Delaunay-like SF network. The long-range links (black solid lines in the left) are exchanged to red ones in the shaded triangles by diagonal flips in the middle and right. The dashed lines are new links from the barycenter, and form new five triangles with contours in the left (The two black solid lines crossed with dashed lines are removed after the second diagonal flip).

Figure 3 shows that the degree distributions follow a power-law with the exponent nearly 3 in RA, log-normal in DT, and power-law with an exponential cutoff in DLSF networks [6]. Such a lognormal distribution has an unimodal shape as similar to one in Erdös-Renyi random networks, and a cutoff is rather natural in real networks [14]. We have found [6, 8] the original RA and DLSF networks without shortcuts are vulnerable because of double constraints of planarity and geographical distances on the linkings in the scale-free structure, but DT networks are not so. Figure 4 shows the examples.

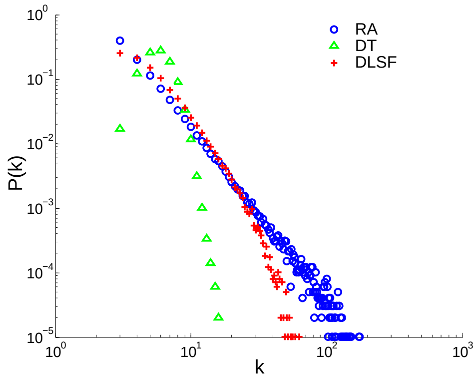


Fig. 3. Degree distribution $P(k)$ for each original network at the total numbers of nodes $N = 1000$ and links $M = 2993$ obtained over 100 realizations.

On the preliminaries, just like overhead superhighways, we add shortcuts between randomly chosen two nodes excluding self-loops and multi-links after constructing the above networks. For adding shortcuts, the routing algorithm can be extended [8] from that in [10]. Note that the added shortcuts contribute to create some higher-order cycles which consists of a long path and the overhead bridge in the majority of triangular cycles. Since we confirm that the degree distributions have only small deviation from the original ones at added shortcut rate up to the amount of 30% of the total links, we can compare the effect of shortcuts on the robustness in the geographical networks under the similar degree distributions.

3 Improved robustness by adding shortcuts

3.1 Simulation results

The fault tolerance and attack vulnerability are known as the typical properties of SF networks [15], which are further affected by geographical constraints. We investigate the tolerance of connectivity in the giant component (GC) of the geographical networks with shortcuts comparing with that of the original ones without shortcuts. The size S of GC and the average size $\langle s \rangle$ of isolated clusters are numerically obtained from ensembles over 100 realizations for each

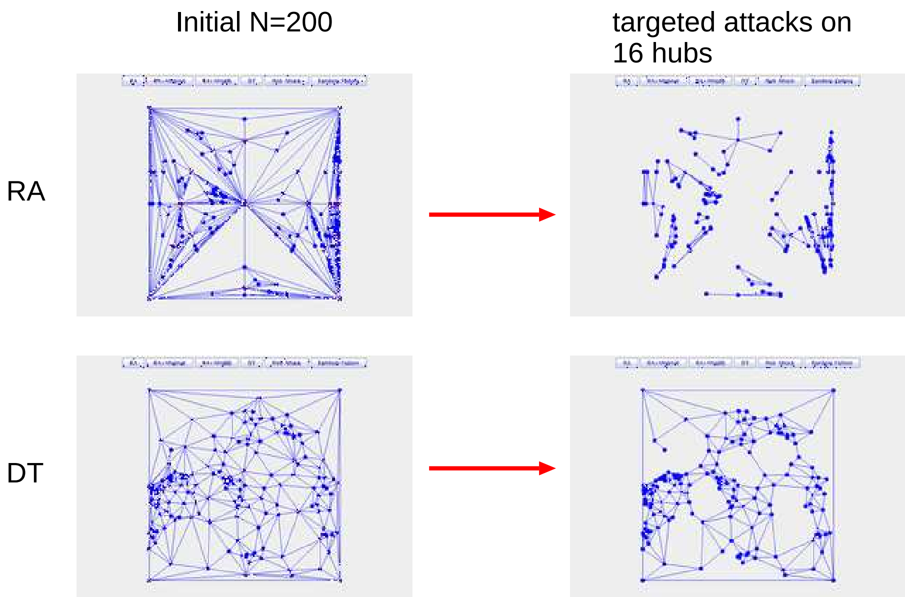
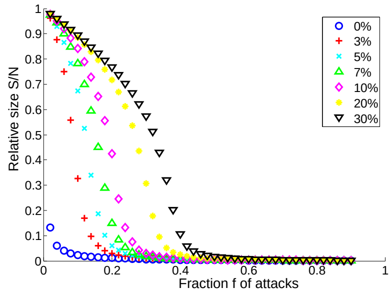


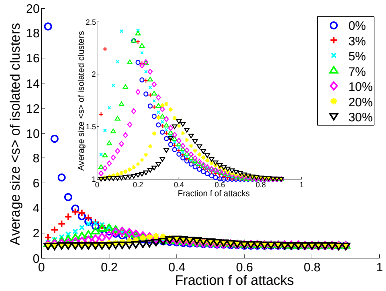
Fig. 4. Examples of isolated clusters which are not able to communicate to each other in RA, while the giant component is still remained in DT.

network model. The critical value of the shortcut rate at the breaking of the GC is estimated as shown in Appendix.

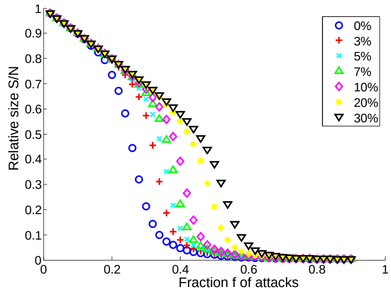
Figure 5 shows the effect of shortcuts on the robustness against the targeted attacks on hubs. The breaking points of GC at the peaks of $\langle s \rangle$ are shifted righter, as the shortcut rate is larger. Around the shortcuts rate 10%, the extremely vulnerable RA and DLSF networks are improved up to the similar level to DTs. It is consistent with the effect in evolving networks with local preferential attachment [16] that the tolerance becomes higher as increasing the cutoff under the same average degree $\langle k \rangle$ and size N. We emphasis that, by adding shortcuts around 10% under almost invariant degree distributions, the robustness against the intentional attacks can be considerably improved up to the similar level to the fully rewired networks by ignoring the geographical constraints [6]. There are also slightly improvements by adding shortcuts for random failures of nodes [8]. We intuitively understand the effect from that the overhead shortcuts consist of the necessary backbone to connect local areas as illustrated in Figure 6.



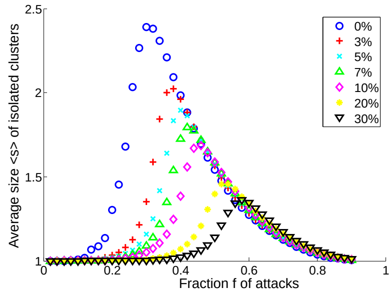
(a) S/N in RA



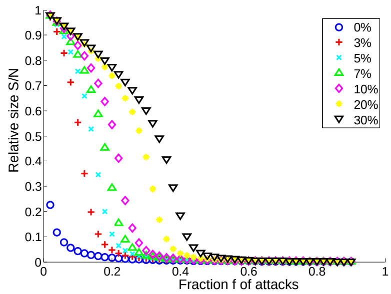
(d) $\langle s \rangle$ in RA



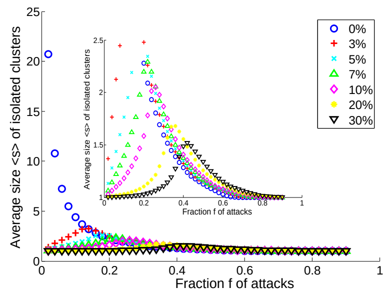
(b) S/N in DT



(e) $\langle s \rangle$ in DT



(c) S/N in DLSF



(f) $\langle s \rangle$ in DLSF

Fig. 5. Relative size S/N of the GC in (a)-(c), and the average size $\langle s \rangle$ of isolated clusters except of the GC in (d)-(f) against intentional attacks. Each shortcut rate is denoted by a different mark in legend. Inset shows the peaks enlarged by other scale of the vertical axis. The robustness is more improved by larger shortcut rate.

3.2 Centrality of superhighways

We investigate how much the shortcut link is used in transport on the shortest distant path routing in order to study necessary backbone recovered from the breaking in the original networks by adding shortcuts. We define the

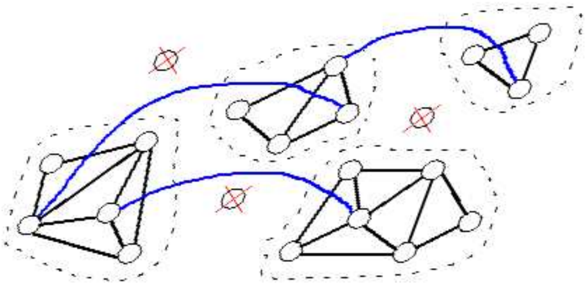
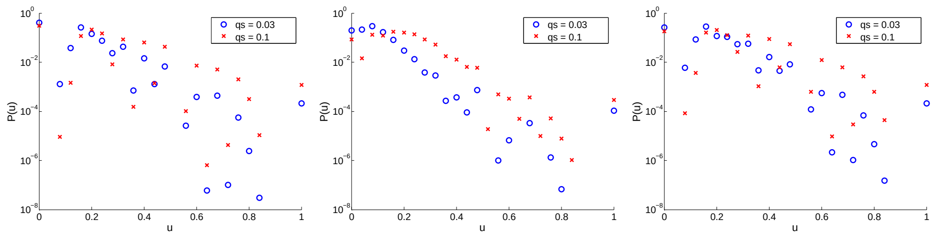


Fig. 6. Backbone of superhighways: The robustness of connectivity is increased by adding shortcuts, since shortcut links (blue) bridge isolated faces by attacks in the original planar network. The mark with (red) cross denotes the removed nodes.

distribution $P(u)$ of superhighway usage $u \stackrel{\text{def}}{=} l_{super}/l_{short}$, where l_{super} is the number of links in a given shortest path of length l_{short} belonging to shortcuts as superhighways, by slightly modifying from a path based on the MST [17] to one on the shortest distance. Figure 7 shows that shortcuts are frequently used on the shortest path in spite of only 3 or 10 % of the total links, and that the distribution $P(u)$ looks like an exponential decaying.

Without loss of generality, we assume both source and terminal are chosen in uniformly random from all nodes at every time step. Thus, the packet generation and receiving seem to be homogeneous with an equal probability for all nodes, however the spatial distribution is remarkably heterogeneous according to the node densities (remember Figure 1). This situation based on human activities is realistic, since the packets are more generated and received as the population is larger in a dense area.



(a) RA (b) DT (c) DLSF

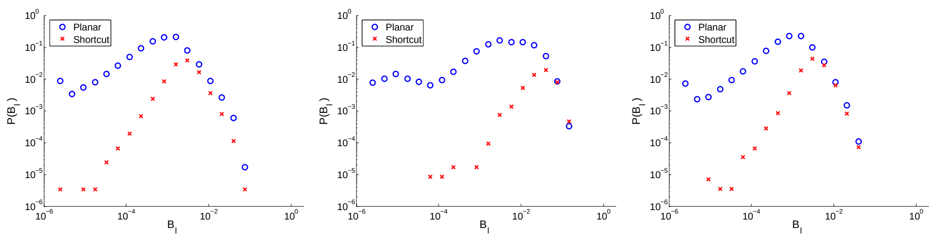
Fig. 7. Distribution $P(u)$ of superhighway usage u obtained over 100 realizations. The average usage $\langle u \rangle$ is 0.1225, 0.0769, 0.1535 at shortcut rate $q_s = 0.03$ (0.1849, 0.1764, 0.2236 at $q_s = 0.1$) in RA, DT, and DLSF networks, respectively.

Let us return to the study of superhighways. For both shortcuts and the other planar links, we consider the normalized betweenness centrality of link l [18],

$$B_l \stackrel{\text{def}}{=} \frac{2}{(N-2)(N-3)} \sum_{k \neq j \neq l1, l2} \frac{b_k^j(l)}{b_k^j},$$

where $l1$ and $l2$ denote the end nodes of l, b_k^j is the number of shortest path between any nodes k and j, and $b_k^j(l)$ is the number of such path passing through the link l. Figures 8 and 9 show the peak for shortcuts has higher centrality than that for planar links. Note that on the shortest path shortcuts act as necessary bridges between isolated clusters, however planar links are used on the path only in a cluster (remember Figure 6). Therefore it is natural to be high centrality in some planar links, while the majority of other planar links have low centralites. We confirm these results are not depended on the initial configuration of networks. Similar results are obtained for the normalized effective betweenness [19] in which $l1$ and $l2$ are included as the source and terminal k, j,

$$\hat{B}_l \stackrel{\text{def}}{=} \frac{2}{N(N-1)} \sum_{k<j} \frac{b_k^j(l)}{b_k^j}.$$

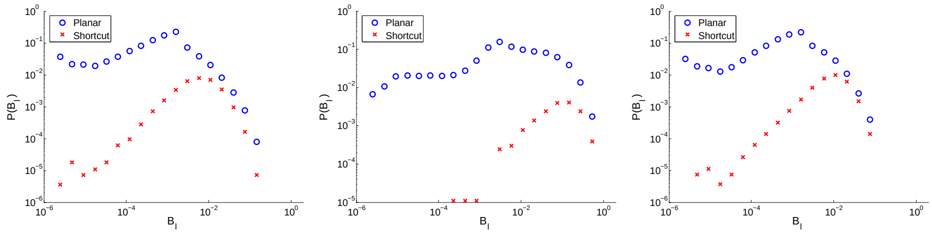


(a) RA (b) DT (c) DLSF

Fig. 8. Distribution $P(B_l)$ of betweeness centrality of link in each geographical network with shortcuts ($q_s = 0.1$) after the intentional attacks on hubs at the breaking of the GC in the original one obtained over 100 realizations.

4 Conclusion

For the prediction of vulnerable connectivity caused by spatial constraints, we have shown the improvement in geographical networks by adding a small fraction of shortcuts between randomly chosen two nodes. In particular, we study a class of planar networks called RA, DT, and DLSF, which are practically



(a) RA (b) DT (c) DLSF

Fig. 9. Distribution $P(B_l)$ of betweeness centrality of link in each geographical network with fewer shortcuts ($q_s = 0.03$) after the intentional attacks on hubs at the breaking of the GC in the original one obtained over 100 realizations.

suitable for wireless communication without interference, growth of network according to the increasing population, and distributed routing algorithm with only local information.

Considering the transport properties, the usage on the shortest path and the centrality of a link have been investigated in comparison with shortcuts and the other planar links. Our results have shown that the shortcut is not only effective to avoid the serious breaking for the intentional attacks on hubs, but also constructs a necessary backbone such as superhighways to bridge isolated clusters. In addition, we point out a realistic situation in the spatially heterogeneous generation and receiving of packets which depend on a naturally emerged node density through the growing of geographical networks. Further studies will be required for the development of more robust and efficient networks on such geographical heterogeneity.

Acknowledgment

The author would like to thank Jun Matsukubo for helping the simulation. This research is supported in part by Grant-in-Aide for Scientific Research in Japan, No.18500049.

Appendix: Percolation analysis for adding random shortcuts

If we know the distribution $P(n)$ of size n of clusters isolated by the intentional attacks in each geographical network without shortcuts, we can estimate the critical value of shortcut rate at emerging the whole connected component by applying the generating function approach [5]. However, it is difficult to analytically obtain it. Thus, we numerically predict the approximative form.

Figures 10(a)(b) show the cluster size distribution that numerically estimated as a power-law $P(n) \sim n^{-\alpha}$, $\alpha \approx 2.1$ in DT and as an exponential $P(n) \sim \exp(\beta n)$, $\beta \approx 0.01$ in DLSF[1] at the breaking point of GC as shown in Figure 5. Based on the estimated $P(n)$, we investigate the critical value of shortcut rate at which the isolated clusters by attacks can be connected with adding shortcuts. The generating function is given by

$$\mathcal{H}_0(x) = \sum_{n=0}^{\infty} P(n) x^n \sum_{m=0}^{\infty} P(m|n)[\mathcal{H}_0(x)]^m, \quad (1)$$

where $P(m|n)$ is the conditional probability that there are exactly m shortcuts emerging from a local cluster of size n [5]. At the shortcut rate q_s, the total number of shortcut links are $q_s \langle k \rangle N/2$, then the conditional probability is given by the combination of m links for the both ends of shortcut links in randomly chosen n nodes from the total N,

$$P(m|n) = {}_{q_s\langle k \rangle N} C_m \left(\frac{n}{N}\right)^m \left(1 - \frac{n}{N}\right)^{q_s \langle k \rangle N - m},$$

By substituting this into Eq. (1) and using binomial expansion with the formula $\lim_{N' \to \infty}(1 + c/N')^{N'} = \exp(c)$, we obtain

$$\begin{aligned} \mathcal{H}_0(x) &= \textstyle\sum_{n=0}^{\infty} P(n) x^n \left[1 + (\mathcal{H}_0(x) - 1)\frac{n}{N}\right]^{q_s \langle k \rangle N} \\ &\approx \textstyle\sum_{n=0}^{\infty} P(n) \left[x \exp(q_s \langle k \rangle (\mathcal{H}_0(x) - 1))\right]^n \\ &= \bar{\mathcal{H}}_0(x \exp(q_s \langle k \rangle (\mathcal{H}_0(x) - 1))). \end{aligned}$$

where we denote another generating fuction for the distribution of cluster sizes in the broken original networks as $\bar{\mathcal{H}}_0(z) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} P(n) z^n$. Thus, the average cluster size connected with adding shortcuts is given by

$$\langle s \rangle = \mathcal{H}_0'(1) = \frac{\bar{\mathcal{H}}_0'(1)}{1 - q_s \langle k \rangle \bar{\mathcal{H}}_0'(1)}.$$

At the divergence of this denominator, the GC becomes dominant in $\langle s \rangle \to \infty$, the critical value is

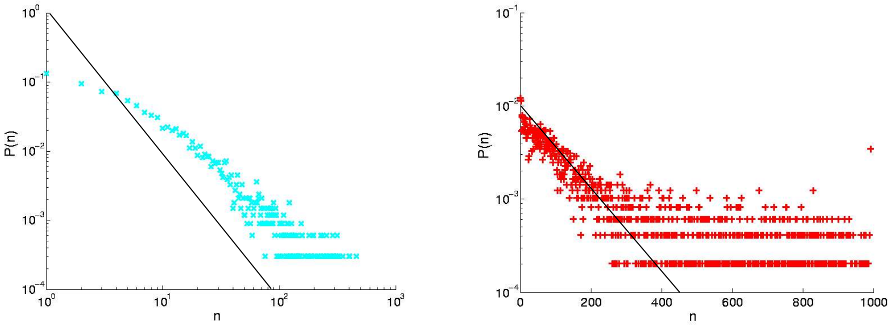
$$q_s = 1/(\langle k \rangle \bar{\mathcal{H}}_0'(1)). \quad (2)$$

DT : power-law distribution of cluster size

From Eq.(2), a power-law distribution $P(n) = C_{DT} n^{-\alpha}$, and

$$\bar{\mathcal{H}}_0'(1) = \sum_n n P(n) \approx C_{DT} \int_1^N n^{1-\alpha} dn \to \frac{C_{DT}}{\alpha - 2},$$

[1] It is omitted in RA, since the function form is unclear with larger fluctuation.



(a) DT (b) DLSF

Fig. 10. Distribution $P(n)$ of cluster size n at the breaking point of the GC in the original networks without shortcuts. The solid lines guide (a) power-law $P(n) \sim n^{-2.1n}$ and (b) exponential $P(n) \sim \exp(-0.01n)$.

we obtain $q_s = \frac{\alpha-2}{\langle k \rangle C_{DT}}$ at $N \to \infty$, where $C_{DT} \approx \alpha - 1$ is the normalization constant.

In the setting $\langle k \rangle = 5.986$ and $\alpha \approx 2.1$, the estimated value $q_s = 0.0157$ coincides with the result that nearly 2% of shortcuts connect the isolated clusters by the 30% of hub attacks at the breaking point of the GC ($f = 0.3$) as shown in Figure 5: jumping from circle to plus point.

DLSF : exponential distribution of cluster size

From Eq.(2), an exponential distribution $P(n) = C_{DL} \exp(-\beta n)$, and

$$\bar{\mathcal{H}}_0'(1) = \sum_n nP(n) \approx C_{DL} \int_1^N n \exp(-\beta n) dn \to \frac{C_{DL}(1+\beta)\exp(-\beta)}{\beta^2},$$

we obtain $q_s = \frac{\beta^2 \exp(\beta)}{\langle k \rangle (1+\beta) C_{DL}}$ at $N \to \infty$, where $C_{DL} \approx \beta \exp(\beta)$ is the normalization canstant.

In the setting $\langle k \rangle = 5.986$ and $\beta \approx 0.01$, the estimated value $q_s = 0.00165$ coincides with the result that nearly 1% of shortcuts connect the isolated clusters by the 3% of hub attacks at the breaking point of the GC ($f = 0.03$) as shown in Figure 5: jumping from circle to plus point.

References

1. S.-H. Yook, H. Jeong, and A.-L. Barabási, *PNAS* **99(21)**,13382, (2002).
2. M.T. Gastner, and M.E.J. Newman, *Eur. Phys. J. B* **49(2)**, 247, (2006).
3. L. Huang, L. Yang, and K. Yang, *Europhys. Lett.* **72(1)**, 144-150, (2005).

4. L. Huang, L. Yang, and K. Yang, *Phys. Rev. E* **75**, 036101, (2007).
5. C. Moore, and M.E.J. Newman, *Phys. Rev. E* **61**, 5678, (2000).
6. Y. Hayashi, and J. Matsukubo, *Phys. Rev. E* **73**, 066113, (2006).
7. K. Imai, *IEICE Trans. on Infor. and Syst.* **83-D(3)**, 428, (2000).
8. Y. Hayashi, and J. Matsukubo, *Physica A* **380**, 552-562, (2007).
9. T.W. Haynes, S.T. Hedetniemi, and P.J. Slater, Fundamentals of Domination in Graphs, Marcel Dekker, Inc., 1998.
10. P. Bose, and P. Morin, *SIAM J. of Computing* **33(4)**, 937, (2004).
11. J.M. Keil, and C.A. Gutwin, *Discrete Compt. Geom.* **7**, 13-28, (1992).
12. J.P.K. Doye, and C.P. Massen, *Phys. Rev. E* **71**, 016128, (2005).
13. T. Zhou, G. Yan, and B.-H. Wang, *Phys. Rev. E* **71**, 046141, (2005).
14. L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, *PNAS* **97(21)**, 11149, (2000).
15. R. Albert, and A.-L. Barabási, *Nature* **406**, 378, (2000).
16. S. Sun, Z. Liu, Z. Chen, and Z. Yuan, *Physica A* **373**, 851-860, (2006).
17. Z. Wu, L.A. Braunstein, S. Havlin, and H.E. Stanley, *Phys. Rev. Lett.* **96**, 148702, (2006).
18. L.C. Freeman, S.P. Borgatti, and D.R. White, *Social Networks* **13**, 141-154, (1991).
19. R. Guimera, A. D.-Guilera, F. V.-Redondo, A. Cabrales, and A. Arenas, *Phys. Rev. Lett.* **89(24)**, 248701, (2002).