A general framework for analyzing long-range degree correlations in complex networks

Yuka Fujiki[1], Taro Takaguchi[2], and Kousuke Yakubo[1]

[1]Department of Applied Physics, Hokkaido University, Sapporo 060-8628, Japan
[2]National Institute of Information and Communications Technology, Tokyo 184-8795, Japan

Keywords Scale-Free Network | Degree-Degree Correlation

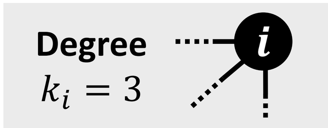
https://arxiv.org/abs/1712.00910

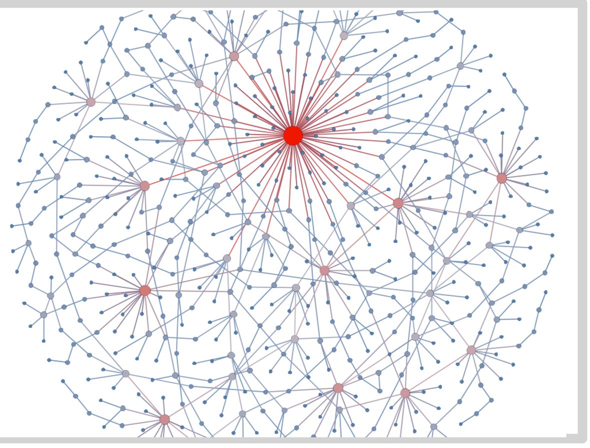
1. Introduction

Degree-degree correlations in real-world networks

Scale-Free Nature of degree distribution

$$P(k) \propto k^{-\gamma}$$



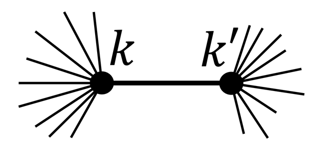


⇨ High degree nodes (**Hubs**) exist and play important roles.

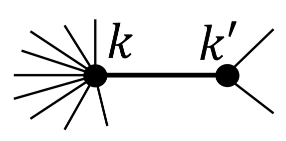
⇨ **Degree-degree correlations** (complexity peculiar to networks)

Previous research tends to focus on adjacent nodes

Nearest neighbor degree correlation (NNDC)



Assortative mixing
Social networks



Disassortative mixing
Biological or Technological networks

Assortativity [M. E. J. Newman (2002)]

$$r = \frac{4\sum_{k,k'} kk'P(k,k') - \left[\sum_{k,k'}(k+k')P(k,k')\right]^2}{2\sum_{k,k'}(k^2+k'^2)P(k,k') - \left[\sum_{k,k'}(k+k')P(k,k')\right]^2}$$

Average degree of nearest neighbor nodes

$$k_{nn}(k) = \sum k' P(k'|k)$$

More generally,

- $P_{nn}(k,k')$: Probability that one end node of a randomly chosen edge has the degree k and the other end node has the degree k'
- $P_{nn}(k'|k)$: Probability that a node adjacent to a randomly chosen node of degree k has the degree k'

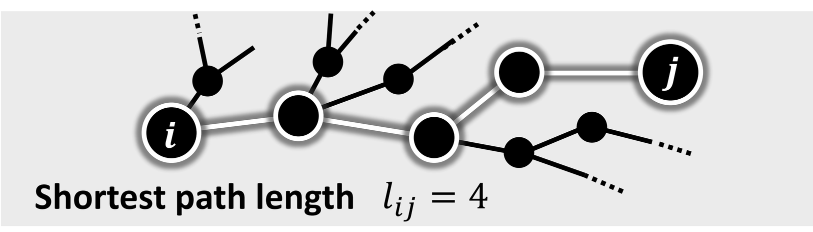
Impact of nearest neighbor degree correlations (NNDCs)

- Resilience for random or targeted attack [M. A. Serrano (2006), A. V. Gol'tsev (2008)]
- Diffusion of information or disease [A.-L. Barabási (2016), C. E. Gross (2006)]
- Synchronization of oscillator [C. E. La Rocca (2011), V. Avalos-Gaytan (2012)]
- Game theory [Z. Rong (2007)]

However

NNDC is not enough to characterize degree-degree correlations.

Degree correlations between nodes **separated by more than one step** (i.e., beyond nearest neighbors)



⇨ **Long-Range Degree Correlation (LRDC)**

NNDC is not enough to explain global properties

- Long-range hub repulsion in fractal networks [Y. Fujiki (2017)]
- Reconstruction of networks by NNDC [C. Orsini (2015)]

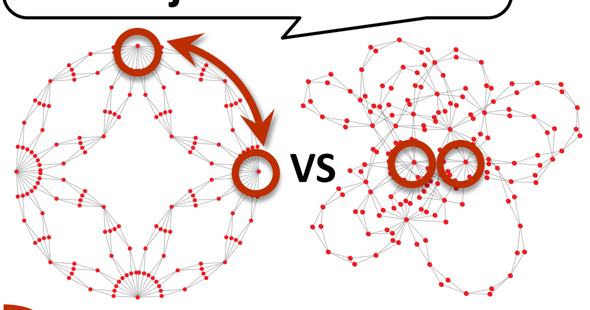
Path lengths between hubs affects on dynamics

- Jamming threshold in communication networks [B. Tadić (2004)]
- Epidemic threshold of the SIS model [M. Boguna (2013)]

Previous proposals for formulating LRDCs

- Long-range assortativity [M. Mayo (2015), A. Arcagni (2017)]
- Fluctuations of the degree along shortest paths [D. Rybski (2010)]
- Two-walks degree assortativity [A. Allen-Perkins (2017)]

☺How far?
☹Adjacent or not?



Specific aspects of LRDCs ≠ **General argument of LRDCs**

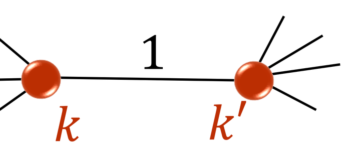
2. Objective

To provide a general description of long-range degree correlations in complex networks

3. Description of LRDC

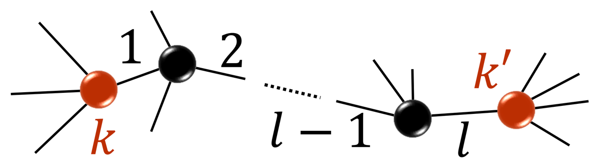
Five probability distributions characterizing LRDC

NNDC described by $P_{nn}(k,k'), P_{nn}(k'|k)$



Two endpoints of an edge

⇔



Two endpoints of an l-length shortest path

Joint distribution

- $P(k,k',l)$: Probability that one node of a randomly chosen node pair has the degree k, the other node has the degree k', and the path length between two nodes is l

Conditional distributions

- $P(k,k'|l)$: Probability that one node of a randomly chosen node pair separated by l from each other has the degree k and the other node has the degree k'
- $P(k',l|k)$: Probability that a randomly chosen node has the degree k' and is separated by l from a node of degree k
- $P(k'|k,l)$: Probability that a node separated by l from a randomly chosen node of degree k has the degree k'
- $P(l|k,k')$: Probability that the path length between randomly chosen two nodes of degrees k and k' is l

Bayes' theorem

$$P(A \cap B) = P(A|B)P(B)$$

$R(l)$ → $P(k,k'|l)$ → $P(k,k',l)$ ← $P(k'|k,l)$ ← $R(l)Q(k|l)$

$P(k',l|k)$ ← $P(k)$; $P(l|k,k')$ ← $P(k)P(k')$

$P(k)$: Degree distribution
$R(l)$: Shortest path length distribution

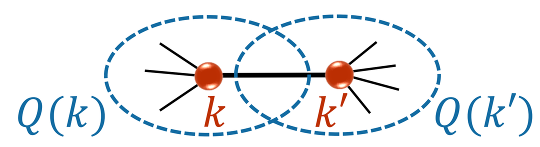
If one of the five probability distributions is given, we can calculate other distributions.

4. Long-range uncorrelated networks (LRUN)

Judgement of existence of LRDCs ⇦ uncorrelated networks

Nearest neighbor uncorrelated networks

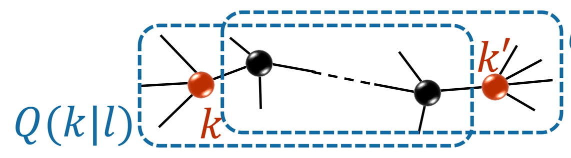
$$P_{nn}(k,k') = Q(k)Q(k')$$



$Q(k|l) = \sum_{k'} P(k,k'|l)$
Probability that one node of a randomly chosen node pair separated by l has the degree k

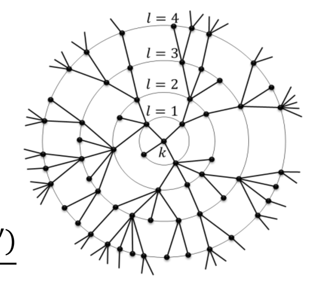
Long-range uncorrelated network (LRUN)

$$P(k,k'|l) = Q(k|l)Q(k'|l) \text{ for } \forall l$$



4.1. Infinite tree-like networks satisfy the condition of LRUN

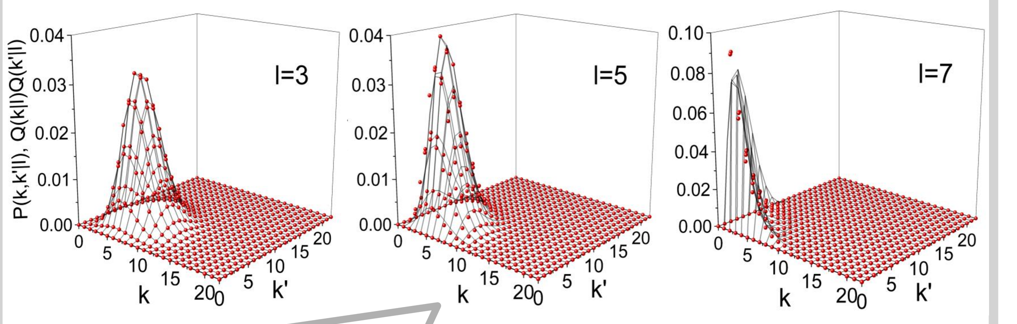
$P(k,k',l) = 0$
$P(k',l|k) = 0$
$P(l|k,k') = 0$
$P(k,k'|l) = \frac{kk'P(k)P(k')}{\langle k\rangle^2}$
$P(k'|k,l) = \sum_{k''} P(k''|k,l-1)P_{nn}(k'|k'') = \frac{k'P(k')}{\langle k\rangle}$



4.2. Can finite networks satisfy that?

It is not easy to answer this question rigorously...

Most plausible candidate for LRUNs = **random networks** (configuration model)



▦ : $P(k,k'|l)$
● : $Q(k|l)Q(k'|l)$

$N = 1{,}000$
$\langle k\rangle = 5.0$
$\langle l\rangle = 4.47$

Confirmation of $P(k,k'|l) = Q(k|l)Q(k'|l)$ for Erdős-Rényi random graphs

$$\begin{cases} P(k,k'|l) \neq Q(k|l)Q(k'|l) & \text{for } l \gg \langle l\rangle \quad \text{(finite-size effect)} \\ P(k,k'|l) = Q(k|l)Q(k'|l) & \text{for } l \ll \langle l\rangle \quad \text{(local tree-like)} \end{cases}$$

From a practical viewpoint **Baseline for comparison** ⇨ **Random networks** (P_0) with the same degree sequence

Calculation of P_0 using the mean-field and local-tree approximations
[S. Melnik and J. P. Gleeson, arXiv:1604.05521 (2016)]

$$P_0(l|k,k') = \rho_l^{kk'} - \rho_{l-1}^{kk'}$$

$\rho_l^{kk'}$: Probability that the distance between randomly chosen two nodes of degrees k and k' is equal to or less than l

$\bar{q}_l^{k'}$: Probability that an adjacent node of a randomly chosen node i lies within l from a k'-degree node j under the condition that i is separated more than l from j

Recursion formula:

$$1-\bar{q}_{l+1}^{k'} = G_1(1-\bar{q}_l^{k'}) - \frac{k'(1-\bar{q}_l^{k'})^{k'-1}}{N\langle k\rangle}$$

Initial state:

$$\rho_0^{kk'} = \frac{\delta_{kk'}}{NP(k)}, \quad \bar{q}_0^{k'} = \frac{k'}{N\langle k\rangle}$$

$$1-\rho_l^{kk'} = [1-\rho_0^{kk'}][1-\bar{q}_{l-1}^{k'}]^k$$

5. Applying to real-world networks

Measure to characterize LRDCs

Extensions of measures for NNDC

$P_{nn}(k,k') = P(k,k'|l=1)$ $\quad$ $P_{nn}(k'|k) = P(k'|k,l=1)$

- Long-range assortativity: [M. Mayo (2015)]

$$r_l = \frac{4\sum_{k,k'} kk'P(k,k'|l) - \left[\sum_{k,k'}(k+k')P(k,k'|l)\right]^2}{2\sum_{k,k'}(k^2+k'^2)P(k,k'|l) - \left[\sum_{k,k'}(k+k')P(k,k'|l)\right]^2}$$

- Average degree of l-th neighbor nodes:

$$k_l(k) = \sum_{k'} k'P(k'|k,l)$$

Other possible indices ....

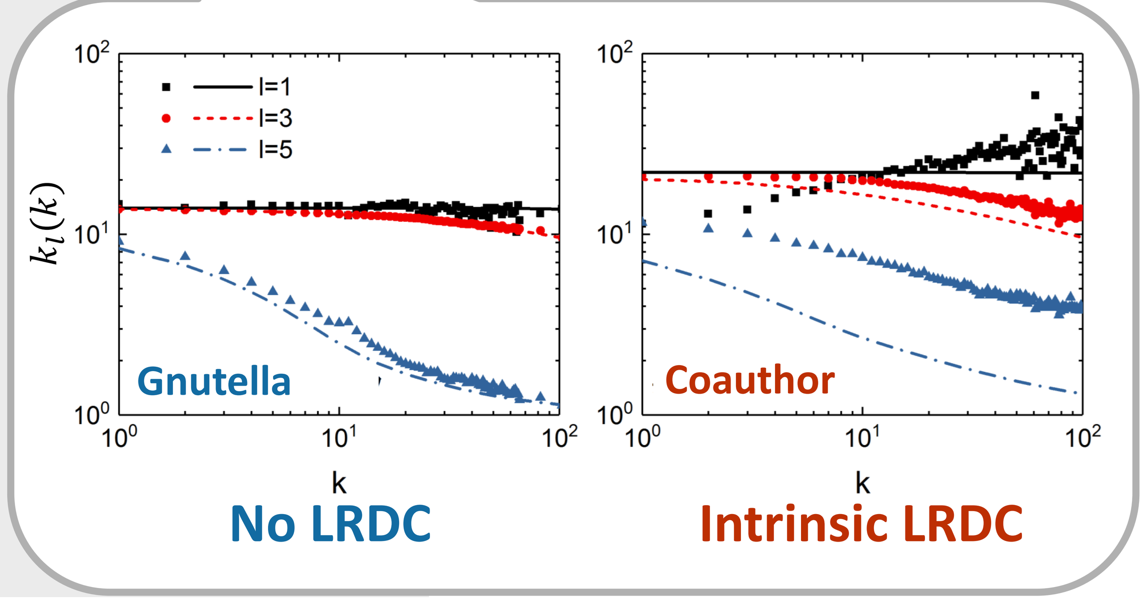
- Rescaled average shortest-path length:

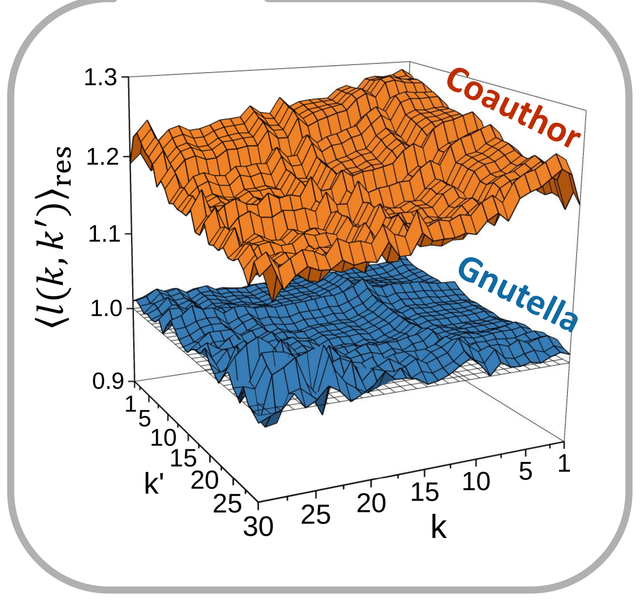
$$\langle l(k,k')\rangle_{res} \equiv \frac{\sum_l lP(l|k,k')}{\sum_l lP_0(l|k,k')}$$

Real-World Networks

Gnutella peer-to-peer network
$N = 10{,}876$ $\quad$ $\langle k\rangle = 7.4$
$E = 39{,}994$ $\quad$ $\langle l\rangle = 4.6$

Coauthorship network
$N = 23{,}133$ $\quad$ $\langle k\rangle = 8.1$
$E = 93{,}439$ $\quad$ $\langle l\rangle = 5.4$





Conclusions

We provided a general framework for analyzing LRDCs.

- To fully describe LRDCs, we introduced fundamental five distributions $P(k,k',l), P(k,k'|l), P(l|k,k'), P(k'|k,l)$, and $P(k',l|k)$. If one of them is given, we can calculate others using Bayes' theorem.
- We adopted random networks as a baseline to judge the existence of LRDCs, instead of LRUNs defined by $P(k,k'|l) = Q(k|l)Q(k'|l)$, and analytically calculated the probability distributions (P_0) for random networks within the mean-field approximation.
- The utility of our argument was demonstrated by applying it to real-world networks.

One can introduce new measures in our framework.