Analysis of Network Heterogeneity by Using Entropy of the Remaining Degree Distribution

Lu Chen, Shin'ichi Arakawa, and Masayuki Murata

Graduate School of Information Science and Technology
Osaka University, Osaka, Japan

Presentation Outline

1. Background and objective
2. Explain the measurement
3. Router-level topologies calculated by the measurement
4. Describing some topological characteristics by changing the value of the measurement through a rewiring process
5. Conclusion and future work

Backgrounds

- Designing the Internet that has adaptability and sustainability against environmental changes is important
  - Adaptability against the failure of network equipment
  - Sustainability against changes of traffic demand
- One of the key properties to focus on is the network heterogeneity
  - "Complex networks display heterogeneous structures from different mechanisms of evolution"[2]

[2] R. Solé and S. Valverde, "Information theory of complex networks: On evolution and architectural constraints," *Complex networks*, vol. 650, pp. 189–207, Aug. 2004.

Goal & Objective

- Goal
  - To design networks that has adaptability and sustainability focusing on the network heterogeneity

⬇ Need a measurement to evaluate the network heterogeneity

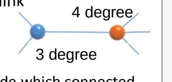
- Objective in this work
  - Confirming mutual information is usable to evaluate the network heterogeneity of topological structure of router-level topologies

Mutual information and Network heterogeneity

- Mutual information
  - The amount of information that can obtain about one random variable X by observing another variable Y
  - $I = H(X) - H_c(X|Y)$
    - $H(X)$: Entropy, $H_c(X|Y)$: Conditional entropy
- Diversity of a topology can be measured
  - Y : a part of the topology
  - X : the rest part of the topology
  - Mutual information is high -> Less diverse
    - Much information can obtain about X by observing Y
  - Mutual information is low -> Diverse
    - A little information can obtain about X by observing Y

Remaining degree distribution as the random variable

- Solé et al.[2] studied complex networks by using remaining degree distribution as the random variable
  - Focus on the relationship of pairs of nodes connected to each other
    - Relationship: degree pattern of those two connected nodes (Number of links connected to a node)
  - Y: degree of a node connected to a randomly selected link
  - X: degree of a node connected to the other end of that link

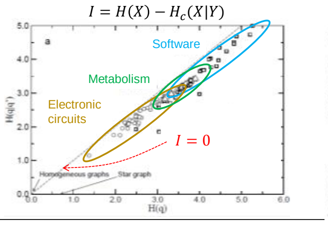


  - Mutual information is high -> Less diverse
    - Much information can obtain about X (the degree of a node which connected to one side of a link) by observing Y (the degree of a node connected to the other side of the link)
  - Mutual information is low -> Diverse
    - A little information can obtain about X by observing Y

Mutual information of complex networks[2]

- Solé et al. calculated mutual information of some complex networks
  - Showing even though I is almost the same, $H(X)$ and $H_c(X|Y)$ is different in some case

$I = H(X) - H_c(X|Y)$



Network type	N	$<k>$	$I(\mathbf{q})$	$H(\mathbf{q})$	$H_c(\mathbf{q}\|\mathbf{q}')$	r
Technological networks						
Software 1	168	2.81	1.19	3.04	1.85	-0.39
Software 2	159	4.19	1.03	3.99	2.97	-0.41
Internet AS	3200	3.56	0.50	4.77	4.27	-0.22
Software 3	1993	5.00	0.30	4.82	4.51	-0.08
Circuit TV	320	3.17	0.23	1.37	1.14	0.010
Circuit EC05	899	4.14	0.15	2.98	2.82	-0.15
Software linux	5285	4.29	0.12	4.47	4.35	-0.06
Powergrid	4941	2.67	0.06	3.01	2.95	0.003
Biological networks						
Silwood park	154	4.75	0.94	4.09	3.14	-0.31
Ythan estuary	134	8.67	0.53	4.74	4.21	-0.24
p53 subnetwork	139	5.09	0.46	4.00	3.54	-0.24
Metabolic map	1173	4.84	0.39	3.58	3.19	-0.17
Neural net (C.elegans)	297	14.5	0.37	5.12	4.74	-0.16
Metabolic map	821	4.76	0.37	3.46	3.09	-0.18
Romanian syntax	5916	5.65	0.31	5.45	5.14	-0.18
Proteome map	1458	2.67	0.24	3.85	3.61	-0.21
Theoretical systems						
Star graph	17	1.88	1.00	1.00	0.00	-1.00
Barabási-Albert	3000	3.98	0.25	4.12	3.85	-0.078
Erdös-Renyi	300	6.82	0.06	3.31	3.25	-0.005
Modular E-R	500	10.3	0.04	3.67	3.62	-0.001

Diversity of Router-level Topology

- I of most of the router-level topologies[11] are higher than that of the model-based ones
  - Router-level topologies are less diverse than model-based ones
  - Regularity comes from technological constraints
- I of Verio is low
  - Verio is more diverse than other router-level topologies
  - This is because Verio grows big with small ISPs so that it contains various kinds of design principles

	Router-level Topologies					Model-based Topologies	
	Telstra	Sprint	AT&T	Level3	Verio	BA	Random
Nodes	329	467	523	623	839	523	523
Links	615	1280	1304	5298	1885	1304	1304
Entropy H	4.24	4.74	4.46	6.04	4.65	4.24	3.22
Conditional Entropy H_c	3.11	3.84	3.58	5.42	4.32	3.98	3.15
Mutual Information I	1.13	0.9	0.88	0.61	0.33	0.26	0.07

Next, we explore the relationship between entropy, conditional entropy and the characteristic of topologies

[11] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with rocketfuel," *IEEE/ACM Transactions on Networking*, vol. 12, pp. 2–16, Feb. 2004.

Detail definition of the measurement

Mutual information of remaining degree $I(\boldsymbol{q})$ [1]:

$$I(\boldsymbol{q}) = H(\boldsymbol{q}) - H_c(\boldsymbol{q})$$

$\boldsymbol{q}(k)$: Remaining Degree Distribution
- Distribution of remaining degree k

k : Remaining Degree
- The number of links leaving the vertex other than the one we arrived along when selected a link

$k = 2$

$H(\boldsymbol{q})$: Entropy
- High when $\boldsymbol{q}(k)$ is heterogeneous

$H(\boldsymbol{q})$ describe heterogeneity of degree distribution

$H_c(\boldsymbol{q})$: Conditional Entropy
- $H_c(\boldsymbol{q})$ is high:
  - Though knowing that a node connected to a selected link has k_1, it is hard to guess k_2 of the other node connected to the other side of the link

$H_c(\boldsymbol{q})$ describe how less biased the combinations of a pair of connected node's remaining degrees are

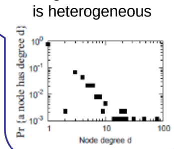
Examples

H is the lowest: All the nodes have the same remaining degree

Ring topology
$I = 0$
$(H = 0, H_c = 0)$

Abilene-inspired topology[7]
$I = 1.02$
$(H = 3.27, H_c = 2.25)$

H is higher than ring: Degree distribution is heterogeneous



H_c is the lowest: When selecting a link, and knowing that one node connected to it has $k_1 = 1$, the other node connected to the other side of the link absolutely has $k_2 = 1$

H_c is low: When knowing that one node connected to a selected link has k_1, the other node connected to the other side of the link has a high probability to have a certain k_2

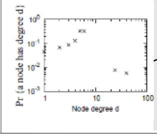
[7] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A first-principles approach to understanding the Internet's router-level topology," *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 3–14, Oct. 2004.

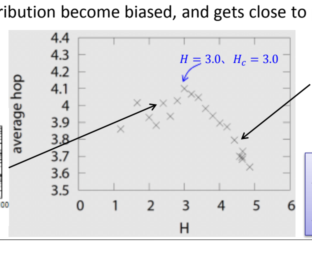
Entropy and the characteristic of topologies

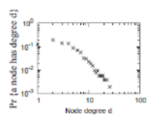
- Generate topologies having different entropy, and compared their average hop distance and degree distribution
- Generating topology has pre-specified H and H_c
  - Minimizing the potential function $U(G)$ by simulated annealing
    - $\boldsymbol{U(G)} = \sqrt{(\boldsymbol{H} - \boldsymbol{H(G)})^2 + (\boldsymbol{H_c} - \boldsymbol{H_c(G)})^2}$
      - $H(G)$ and $H_c(G)$ are calculated by the topology G generated in the optimizing search process
  - Initial topology
    - Obtained by BA model (same number of nodes and links with AT&T)
  - Changing method
    - Random rewiring

Entropy and average hop

- Average hops of topologies obtained by setting H , H_c as $H = H_c$ from 1 to 5
  - $U(G)$ converge to approximately zero
- When H increases higher than 3, the average hop distance decreases
  - Degree distribution become biased, and gets close to power-law around $H = 4$



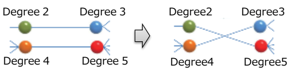




Because router-level topologies obey power-law, next, we compare topologies having high H

Mutual Information and the Characteristic of Topologies

- Generating topologies having different H_c, but having the same degree distribution, and compared their diversity
  - Topologies having the same degree distribution has the same H
  - under the same H, changing H_c is equal to changing I ($I = H - H_c$)
- Generating topology has pre-specified I
  - Minimizing the potential function $U^I(G)$ by simulated annealing
    - $\boldsymbol{U^I(G) = |I - I(G)|}$
      - $I(G)$ is calculated by the topology G generated in the optimizing search process
  - Initial topology
    - Obtained by BA model (same number of nodes and links with AT&T)
  - Changing method
    - Random rewiring that leaves the degree distribution unchanged[14]

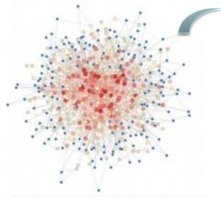


[14] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic topology analysis and generation using degree correlations," in *ACM SIGCOMM Computer Communication Review*, vol. 36, pp. 135–146, Oct. 2006.

Topology $T_{I_{min}}$ with the minimum mutual information

- Topology which has the minimum mutual information under a certain degree distribution

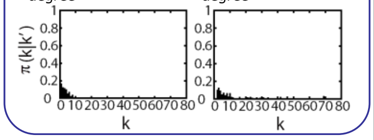
Setting pre- specified $I = 0$



$I(G) = 0.12$
$(H = 4.24、H_c = 4.13)$

The topology is diverse

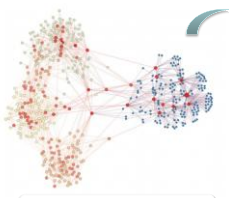
Distribution of remaining degree of nodes connected to the node with k'



Topology $T_{I_{max}}$ with the minimum mutual information

- Topology which has the maximum mutual information under a certain degree distribution

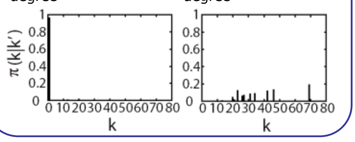
Setting pre- specified I =3



$I(G) = 2.70$
$(H = 4.24、H_c = 1.54)$

The topology is less diverse

Distribution of remaining degree of nodes connected to the node with k'



Conclusion and Future Work

- Conclusion
  - Investigating the network heterogeneity of router-level topologies by using mutual information
    - Router-level topologies have higher mutual information than model-based topologies
  - Generating topologies with different mutual information
    - When the distribution is the same
      - Topology is diverse when mutual information is high
      - Topology has regularity when mutual information is low
- Future work
  - Evaluate network performance of topologies with different mutual information
  - Apply this measure to designing information network that has adaptability and sustainability against environment changes