

Model of Complex Networks based on Citation Dynamics

Lovro Šubelj & Marko Bajec

University of Ljubljana
Faculty of Computer and Information Science

LSNA '13

Introduction

- Real-world networks are scale-free, small-world etc.
- Social networks are **degree assortative**. (Newman and Park, 2003)
↪ Properties captured by many models in the literature.
- However, non-social networks are **degree disassortative**!

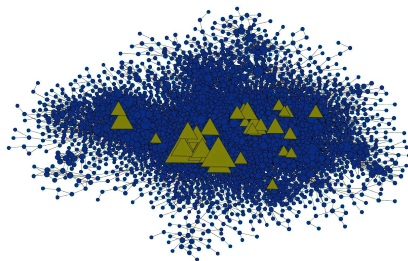


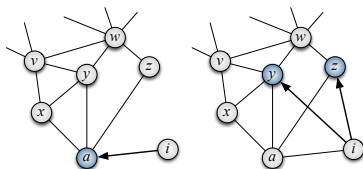
Figure: Part of *Cora* citation network with highlighted hubs.

For simplicity, we consider only undirected networks.

Forest Fire model (Leskovec et al., 2007)

Let p be the *burning probability*.

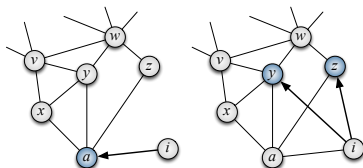
- ① i chooses an ambassador a and links to it;
- ② i selects $x_p \sim G(\frac{p}{1-p})$ neighbors a_1, \dots, a_{x_p} and links to them;
- ③ a_1, \dots, a_{x_p} are taken as the ambassadors of i .



- Networks are scale-free, small-world, **degree assortative** etc.
- Natural interpretation for citation networks!

Author citation dynamics

- ① author chooses a paper (i.e., ambassador) and cites it;
- ② author selects some of its references and cites them;
- ③ the latter are taken as the ambassadors.

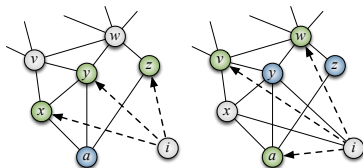


- Assumption \rightarrow *authors read all papers they cite (and vice-versa).*
- Only $\approx 20\%$ of cited papers are read. (Simkin and Roychowdhury, 2003)
- Authors read or cite papers due to **two (independent) processes!**

Citation model _(our)

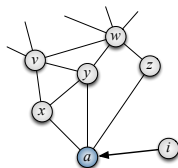
Let q be the *linking probability*.

- ① i chooses an ambassador a ;
- ② i selects $x_p \sim G(\frac{p}{1-p})$ neighbors a_1, \dots, a_{x_p} ;
 i selects $x_q \sim G(\frac{q}{1-q})$ neighbors and links to them;
- ③ a_1, \dots, a_{x_p} are taken as the ambassadors of i .

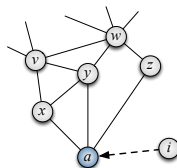
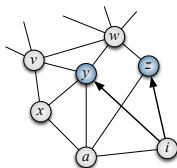


- Networks are scale-free, small-world, **degree disassortative** etc.
- Nodes do not (necessarily) link to their ambassadors!

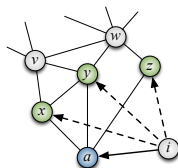
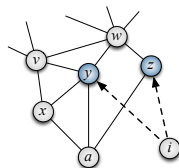
Alternative models



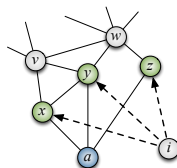
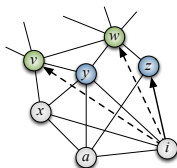
Forest Fire (Leskovec et al., 2007)



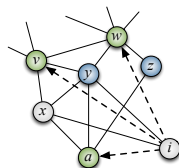
Butterfly (McGlohon et al., 2008)



Copying (Krapivsky and Redner, 2005)



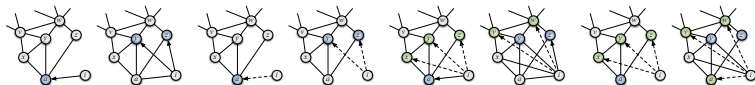
Citation model (our)



Analysis of the models

S , T are the ambassadors and linked nodes.

- **Forest Fire** model: $S = T$
- **Butterfly** model: $S \supseteq T$
- **Copying** model: $S \subseteq T$
- **Citation** model: S, T arbitrary

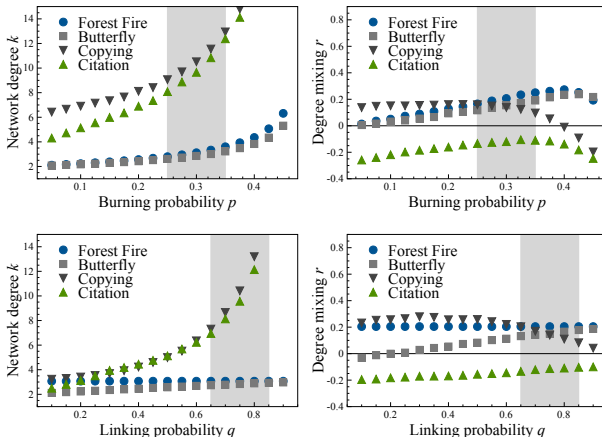


Why degree disassortativity?

- Linking to the ambassadors increases assortativity.
- Absence of such a process prevents assortativity. (Newman and Park, 2003)
- Heterogeneous networks are disassortative. (Johnson et al., 2010)

Comparison of the models (k & r)

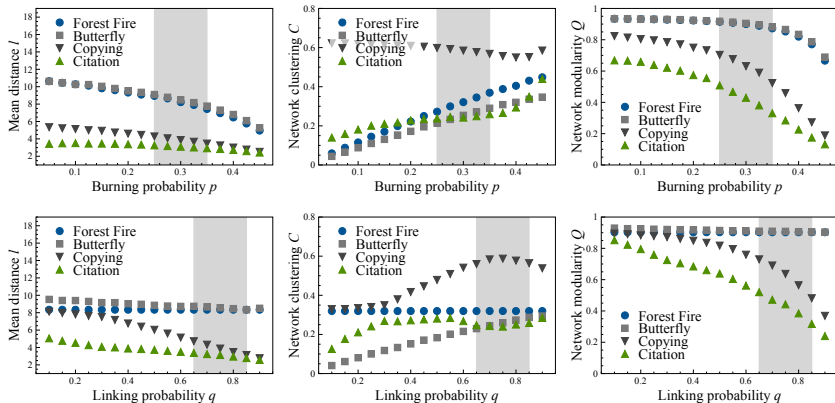
Only Citation model gives **degree disassortative** networks (i.e., $r < 0$).



Shaded regions show most likely parameter values. (Laurienti et al., 2011)

Comparison of the models (I , C & Q)

All models give **(scale-free)** **small-world** networks with high **modularity**.

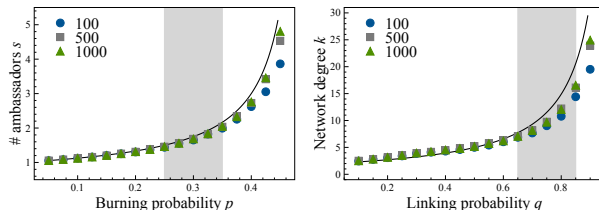


Parameter estimation

s is the number of ambassadors, $s = |S|$.

$$s \leq \frac{1-p}{1-2p} \text{ and } k \leq \frac{2qs}{1-q-(1-q)^{s+1}}$$

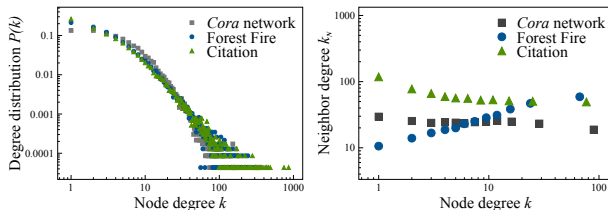
For a given k and fixed q , the system can be solved for p .



Cora citation network

	p	q	n	m	k	r
Cora network			23166	89157	7.697	-0.055
Forest Fire	0.46	-	23166	88828	7.669	0.211
Citation	0.37	0.59	23166	89888	7.760	-0.047

Percentage of papers considered is **66%** (# references just 3.85)!

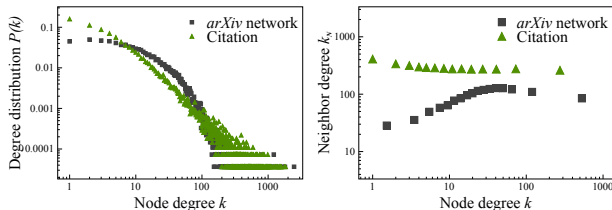


For other network properties see paper and (Šubelj and Bajec, 2012).

arXiv citation network

	p	q	n	m	k	r
arXiv network			27400	352021	25.695	-0.030
Citation	0.46	0.67	27400	350699	25.598	-0.068

Percentage of papers considered is **49%** (# references is 12.85)!



Conclusions

- Model for **citation networks** with most common properties.
- (Non-social) degree non-assortative networks →
nodes must not link to their ambassadors!
- Networks also show **dichotomous mixing**. (Hao and Li, 2011)

Future work:

- extension to directed networks,
- network traversal (isolated nodes),
- analyses on reliable data (e.g., *WoS*).

Questions & comments

`lovro.subelj@fri.uni-lj.si`

`http://lovro.lpt.fri.uni-lj.si/`

- D. Hao and C. Li. The dichotomy in degree correlation of biological networks. *PLoS ONE*, 6 (12):e28322, 2011. doi: 10.1371/journal.pone.0028322.
- S. Johnson, J. J. Torres, J. Marro, and M. A. Muñoz. Entropic origin of disassortativity in complex networks. *Phys. Rev. Lett.*, 104(10):108702, 2010. doi: 10.1103/PhysRevLett.104.108702.
- P. L. Krapivsky and S. Redner. Network growth by copying. *Phys. Rev. E*, 71(3):036118, 2005. doi: 10.1103/PhysRevE.71.036118.
- P. J. Laurienti, K. E. Joyce, Q. K. Telesford, J. H. Burdette, and S. Hayasaka. Universal fractal scaling of self-organized networks. *Physica A*, 390(20):3608–3613, 2011. doi: 10.1016/j.physa.2011.05.011.
- J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):1–41, 2007.
- M. McGlohon, L. Akoglu, and C. Faloutsos. Weighted graphs and disconnected components: Patterns and a generator. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 524–532, New York, NY, USA, 2008.
- M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 68(3):036122, 2003. doi: 10.1103/PhysRevE.68.036122.
- M. V. Simkin and V. P. Roychowdhury. Read before you cite! *Compl. Syst.*, 14:269–274, 2003.
- L. Šubelj and M. Bajec. Clustering assortativity, communities and functional modules in real-world networks. *e-print arXiv:12082518v1*, pages 1–21, 2012.
- L. Šubelj and M. Bajec. Model of complex networks based on citation dynamics. In *Proceedings of the WWW Workshop on Large Scale Network Analysis*, page 4, Rio de Janeiro, Brazil, 2013.