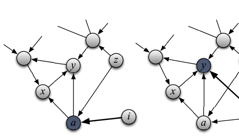
Who reads and who cites? Unveiling author citation dynamics by modeling citation networks

Lovro Šubelj, Slavko Žitnik, Marko Bajec
University of Ljubljana, Faculty of Computer and Information Science
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
{lovro.subelj,slavko.zitnik,marko.bajec}@fri.uni-lj.si

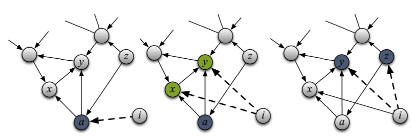
Networks are the simplest representation of complex systems of interacting parts. Despite seemingly plain form real-world networks reveal characteristic structural properties that are absent from regular or random systems. Scale-free degree distributions, small-world structure (i.e., high clustering and short paths), degree mixing (i.e., degree correlations at links' ends) and communities (i.e., densely linked groups) are perhaps the most widely analyzed properties in the literature. Note that communities imply assortative (i.e, positively correlated) mixing by degree that is most notably pronounced in social networks. The above are also the properties captured by many random graph models [1].

However, non-social networks greatly deviate from this picture. Biological and technological networks are in fact degree disassortative (i.e., negatively correlated), while different information networks often reveal no clear degree mixing and lower clustering (e.g., citation networks). We here present an evolving random graph model denoted Citation model [4, 3]. Each newly added node explores the network using the burning process of [1], while links of the visited nodes are copied independently of the latter (Fig. 1). The model generates scale-free networks with arbitrary degree mixing and clustering, and very well reproduces the structure of over 60 years of *Web of Science* data (Fig. 2).

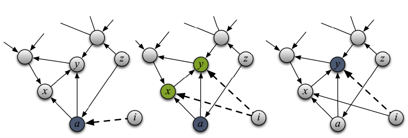
The model also has a natural interpretation for citation networks. The process imitates an author of a paper including references into bibliography (i.e., citation dynamics). Author first reads a related paper, or selects the paper that triggered the research, and cites it and/or some of its references. Author then considers its bibliography for related papers. Some of these are taken for further reading, while the author continues as before.



(a) Forest fire [1]



(b) Citation (our)



(c) Random walk [5]

Figure 1: (a) In Forest fire model [1], new node i chooses an ambassador a and links to it (solid lines). Next, some of its neighbors are taken as the ambassadors by following the in-/out-links (y and z). (b) In our Citation model [4, 3], i links to a with a certain probability (dashed lines), while also to some of its neighbors by following the out-links (x and y). (c) In Random walk model [5], i explores the graph in a random walk fashion.

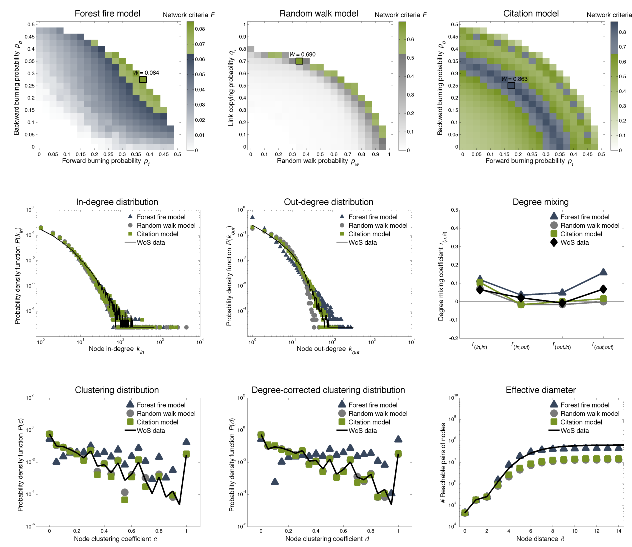


Figure 2: (top) Random graph parameter estimation for citation network of WoS category *Information Science & Library Science* (mind different color scales). (middle) Degree distributions and mixing and (bottom) clustering distributions of WoS citation network and graphs generated with random models.

Based on the above we confirm a 10-year standing claim that 80% of cited papers are never read [2] (due to misprints in bibliographies). In fact, ≈ 85% of references are only copied from other papers for WoS citation data under study. However, previous work failed to recognize that a paper is read for every two cited, while the probability of citing a read paper is merely ≈ 30%. Also, around 30% of cited papers are discovered through paper references, 40% through services providing paper citations and the rest through other sources (e.g., journals, conferences).

Authors thank Thomson Reuters for providing citation data. The work has been supported in part by the Slovenian Research Agency Program No. P2-0359, by the Slovenian Ministry of Education, Science and Sport Grant No. 430-168/2013/91, and by the European Union, European Social Fund.

[1] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):1–41, 2007.

[2] M. V. Simkin and V. P. Roychowdhury. Read before you cite! *Compl. Syst.*, 14:269–274, 2003.

[3] L. Šubelj and M. Bajec. Model of complex networks based on citation dynamics. In *Proceedings of the WWW Workshop on Large Scale Network Analysis*, pages 527–530, Rio de Janeiro, Brazil, 2013.

[4] L. Šubelj, D. Fiala, S. Žitnik, and M. Bajec. Modeling citation network topology. *in preparation*, 2014.

[5] A. Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, 67(5):056104, 2003.