

Modeliranje omrežij citiranj in bibliometrika

Lovro Šubelj

Univerza v Ljubljani,
Fakulteta za računalništvo in informatiko

9.7.2014

Družbena & ostala omrežja

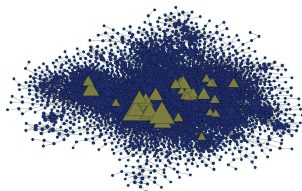
Omrežja so brezlestvična, malega sveta, se gostijo & krčijo.

Nakopičenost C & mešanje r (korelacije stopenj): Watts & Strogatz (1998); Newman (2002)

Družbena \rightarrow *visok C & asortativnost $r > 0$* Newman & Park (2003)

Druga \rightarrow *nižji C & disasortativnost $r < 0$* Šubelj & Bajec (2012)

Citiranja \rightarrow *nižji C & mešanje $r \approx 0$*

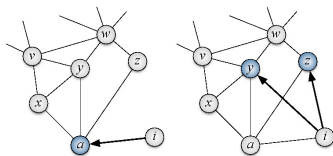


Model gorečega gozda

Zaporedno dodajanje: Leskovec et al. (2007)

- (1) vozlišče izbere ambasadorja a & tvori povezavo z a
- (2) vozlišče izbere sosede a_1, \dots, a_{x_p} & tvori povezave z a_i
- (3) a_1, \dots, a_{x_p} so novi ambasadorji

p je verjetnost gorenja & $x_p \sim G(\frac{p}{1-p})$.

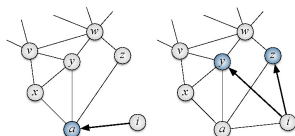


Omrežja so brezlestvična, malega sveta & *asortativna* $r > 0$.

Model omrežij citiranja

Dinamika avtorjev:

- (1) avtor prebere prispevek a & ga citira
- (2) avtor izbere reference a_1, \dots, a_{x_p} & jih citira
- (3) a_1, \dots, a_{x_p} izbrane za nadaljnje branje



Avtorji preberejo vse citirane prispevke...

$\approx 80\%$ citiranih prispevkov ni prebranih. Simkin & Roychowdhury (2003)

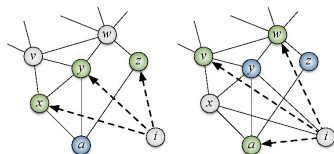
Branje & citiranje prispevkov sta dva procesa!

Model citiranja

Zaporedno dodajanje: Šubelj & Bajec (2013)

- (1) vozlišče izbere ambasadorja a
- (2) vozlišče izbere sosede a_1, \dots, a_{x_p}
vozlišče izbere sosede l_1, \dots, l_{x_q} & tvori povezave z l_i
- (3) a_1, \dots, a_{x_p} so novi ambasadorji

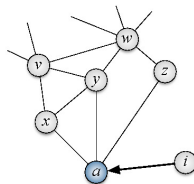
q je verjetnost povezovanja & $x_q \sim G(\frac{q}{1-q})$.



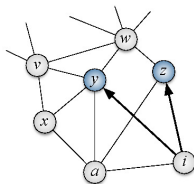
Omrežja so brezlestvična, malega sveta & (dis)asortativna $r \approx 0$.

Analiza alternativnih modelov

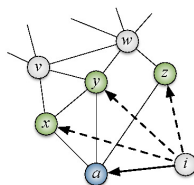
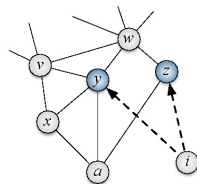
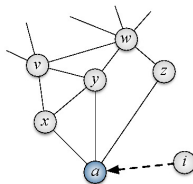
Naj bo A množica ambasadorjev & L množica povezanih vozlišč.



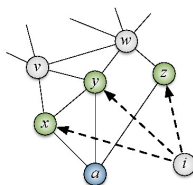
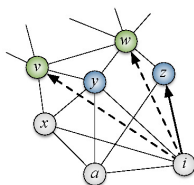
Goreči gozd ($A = L$) Leskovec et al. (2007)



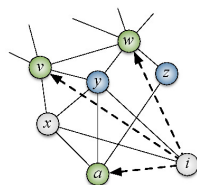
Metuljček ($A \supseteq L$) McGlohon et al. (2008)



Kopiranja ($A \subseteq L$) Krapivsky & Redner (2005)



Citiranja (poljubno) Šubelj & Bajec (2013)

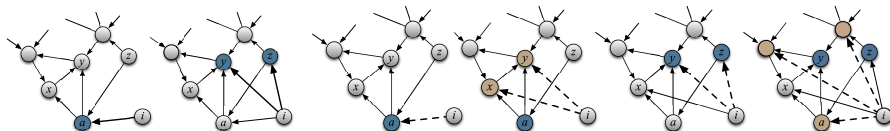


Usmerjen model citiranja

Razširitev na usmerjena omrežja:

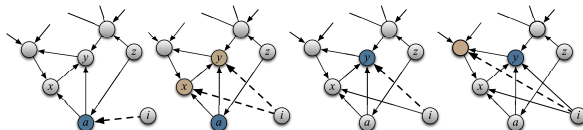
gorenje \rightarrow verjetnosti p_{fwd} & p_{bck}

povezovanje \rightarrow verjetnosti q_{amb} & q_{lnk}



Goreči gozd Leskovec et al. (2007)

Citiranja Šubelj et al. (2014)



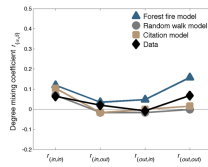
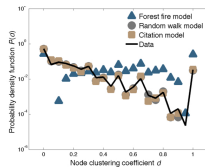
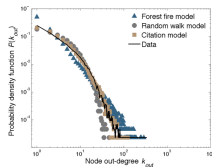
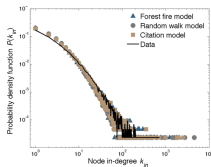
Naključni sprehod Vázquez (2003)

Primerjava modelov citiranj

Dobljena omrežja so brezlestvična & malega sveta.

Model	k	C	D	α_{in}	α_{out}	$r_{(in,in)}$	$r_{(in,out)}$	$r_{(out,in)}$	$r_{(out,out)}$	\mathcal{F}
WoS ILS (1945–)	8.05	0.11	0.14	2.47	3.35	0.06	0.02	−0.01	0.07	1.000
Citiranja (naš)	7.97	0.11	0.13	2.58	3.31	0.12	−0.01	0.01	0.01	0.863
Naključni sprehod	7.91	0.11	0.13	2.61	4.68	0.07	−0.01	−0.02	0.00	0.695
Naključni avtor	7.48	0.21	0.26	2.36	3.26	0.16	−0.02	0.08	0.15	0.400
Goreči gozd	7.88	0.40	0.52	2.51	2.45	0.13	0.04	0.06	0.17	0.087

Le omrežja modela citiranj primerljiva omrežjem citiranj!



Bibliometrika

Citiranih prispevkov: $k/2$.

Prebranih prispevkov: $s = \left(1 - \frac{p_{fwd}}{1-p_{fwd}} - \frac{p_{bck}}{1-p_{bck}}\right)^{-1}$.

% Prebran članek tudi citiran: q_{amb} .

Citatov kopiranih iz drugih prispevkov: $s \cdot \frac{q_{lnk}}{1-q_{lnk}}$.

Razmerje med citati & storitvami: $\frac{p_{fwd}}{1-p_{fwd}} : \frac{p_{bck}}{1-p_{bck}}$.

WoS baza	Citiranje		Branje		Pridobivanje		
	# Citati	% Kopiran	# Branje	% Citiran	% Citati	% Storitve	% Drugo
ILS (1945–)	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
TM (1954–)	2.93	79.7%	1.47	45.2%	74.7%	0.5%	24.9%
AI (1964–)	4.52	87.3%	1.47	40.9%	25.8%	47.6%	26.6%
SE (1954–)	2.78	81.5%	1.58	36.4%	68.8%	2.0%	29.2%
CYB (1961–)	2.18	69.6%	1.59	43.2%	24.5%	37.8%	37.6%

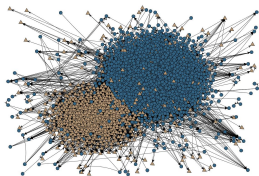
Bibliometrika (II)

WoS ILS	Citiranje		Branje		Pridobivanje		
	# Citati	% Kopiran	# Branje	% Citiran	% Citati	% Storitve	% Drugo
1945–2013	3.98	86.1%	2.14	27.9%	29.2%	41.0%	29.8%
1970–1980	2.23	52.1%	3.39	33.5%	41.4%	0.0%	58.5%
1980–1990	2.62	65.1%	2.96	33.0%	48.3%	1.1%	50.6%
1990–2000	3.42	81.6%	2.38	29.0%	40.3%	23.2%	36.5%
2000–2010	5.06	83.6%	2.90	32.2%	40.7%	27.5%	31.7%

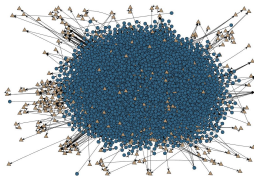
WoS TM	Citiranje		Branje		Pridobivanje		
	# Citati	% Kopiran	# Branje	% Citiran	% Citati	% Storitve	% Drugo
1954–2013	2.93	79.7%	1.47	45.2%	74.7%	0.5%	24.9%
1970–1980	1.99	33.4%	2.27	60.8%	35.4%	24.7%	39.9%
1980–1990	2.52	63.8%	1.76	56.7%	55.1%	10.9%	34.0%
1990–2000	2.93	76.5%	1.61	47.7%	58.8%	12.0%	29.2%
2000–2010	2.96	81.0%	1.45	43.0%	70.8%	2.9%	26.3%

Bibliometrika (III)

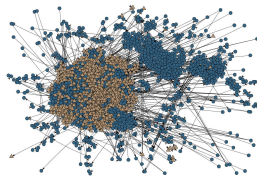
WoS revija	Citiranje		Branje		Pridobivanje		
	# Citati	% Kopiran	# Branje	% Citiran	% Citati	% Storitve	% Drugo
<i>Scientometrics</i>	3.83	87.6%	1.40	38.5%	54.3%	23.4%	22.3%
<i>Expert Syst. Appl.</i>	2.12	81.0%	1.67	25.9%	34.7%	27.8%	37.6%
<i>Commun. ACM</i>	1.32	46.5%	1.70	42.0%	8.7%	38.5%	52.8%



Scientometrics



Expert Syst. Appl.



Commun. ACM

- P. L. Krapivsky & S. Redner. Network growth by copying. *Phys. Rev. E*, 71(3):036118, 2005.
- J. Leskovec, J. Kleinberg, & C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):1–41, 2007.
- M. McGlohon, L. Akoglu, & 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. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20):208701, 2002.
- M. E. J. Newman & J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 68(3):036122, 2003.
- M. V. Simkin & V. P. Roychowdhury. Read before you cite! *Compl. Syst.*, 14:269–274, 2003.
- L. Šubelj & M. Bajec. Ubiquitousness of link-density and link-pattern communities in real-world networks. *Eur. Phys. J. B*, 85(1):32, 2012.
- L. Šubelj & 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.
- L. Šubelj, D. Fiala, S. Žitnik, & M. Bajec. Modeling citation network topology: A natural graph model and applications to citation theory. *in preparation*, 2014.
- A. Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E*, 67(5):056104, 2003.
- D. J. Watts & S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.