

Threshold selection for pseudo-bimodal networks of retweets via different metrics of network centrality

The Fifth International Conference on Network Analysis NET 2015
Laboratory of Algorithms and Techniques for Network Analysis (LATNA)

Alexander Semenov, Igor Zakhlebin and Alexander Tolmach

May 18 2015

Where am I from?



The screenshot shows the website for the International Laboratory for Applied Network Research (ANR). The page features a navigation menu on the left, a main content area with news, and a right sidebar with publications.

Navigation Menu (Left):

- О Лаборатории
- Сотрудники
- Семинары
- Летние школы
- Ресурсы
- Конференции
- Научный руководитель лаборатории
- Цели и задачи
- Заведующий лабораторией

Header: → Научные подразделения → Международная лаборатория прикладного сетевого анализа

Международная лаборатория прикладного сетевого анализа

Новости

10 2014
Профессор Вассерман примет участие в качестве приглашенного спикера на симпозиуме "Multivariate Techniques for the Analysis of Network Data"
Профессор Вассерман примет участие в качестве приглашенного спикера на симпозиуме "Multivariate Techniques for the Analysis of Network Data", который пройдет 13 ноября 2014 года в Салерно, Италия.

24 2014
Мы в новостях Университета Индианы
IU network researcher Stanley Wasserman in collaboration with faculty from the National Research University Higher School of Economics (NRU HSE) in Russia received a \$1.5 million grant to set up an International Laboratory for Applied Network Research at the Russian university.

В Международной лаборатории прикладного сетевого

ПУБЛИКАЦИИ

Книга

[Электроэнергетика России: проблемы выбора модели развития: анализ, докл. к XV Апр. междунар. науч. конф. по проблемам развития экономики и общества, Москва, 1-4 апр. 2014 г.](#)

Баркин О., Волкова И. О., Кожуховский И. и др.
М.: Идентификационный дом НИУ ВШЭ, 2014.

Статья

[Measuring influence of internationalized universities on smart city development in terms of human capital and urban](#)

<http://anr.hse.ru>

Background: Meetings on the 24.11.2011

Opposition Meeting at Prospekt Sakharova



Pro-Government Meeting at Poklonnaya Gora



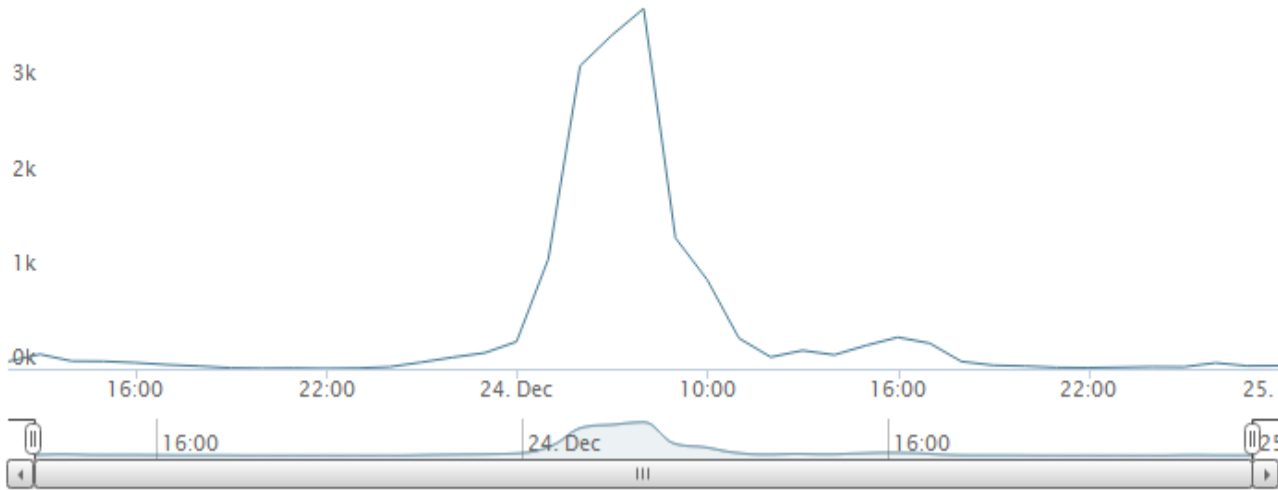
VS



#24дек

Data

Timeframe: December 23, 2011 (07:59) – December 25, 2011 (00:30)



Interactions Received
17196

Filter:
twitter.text contains "24дек"

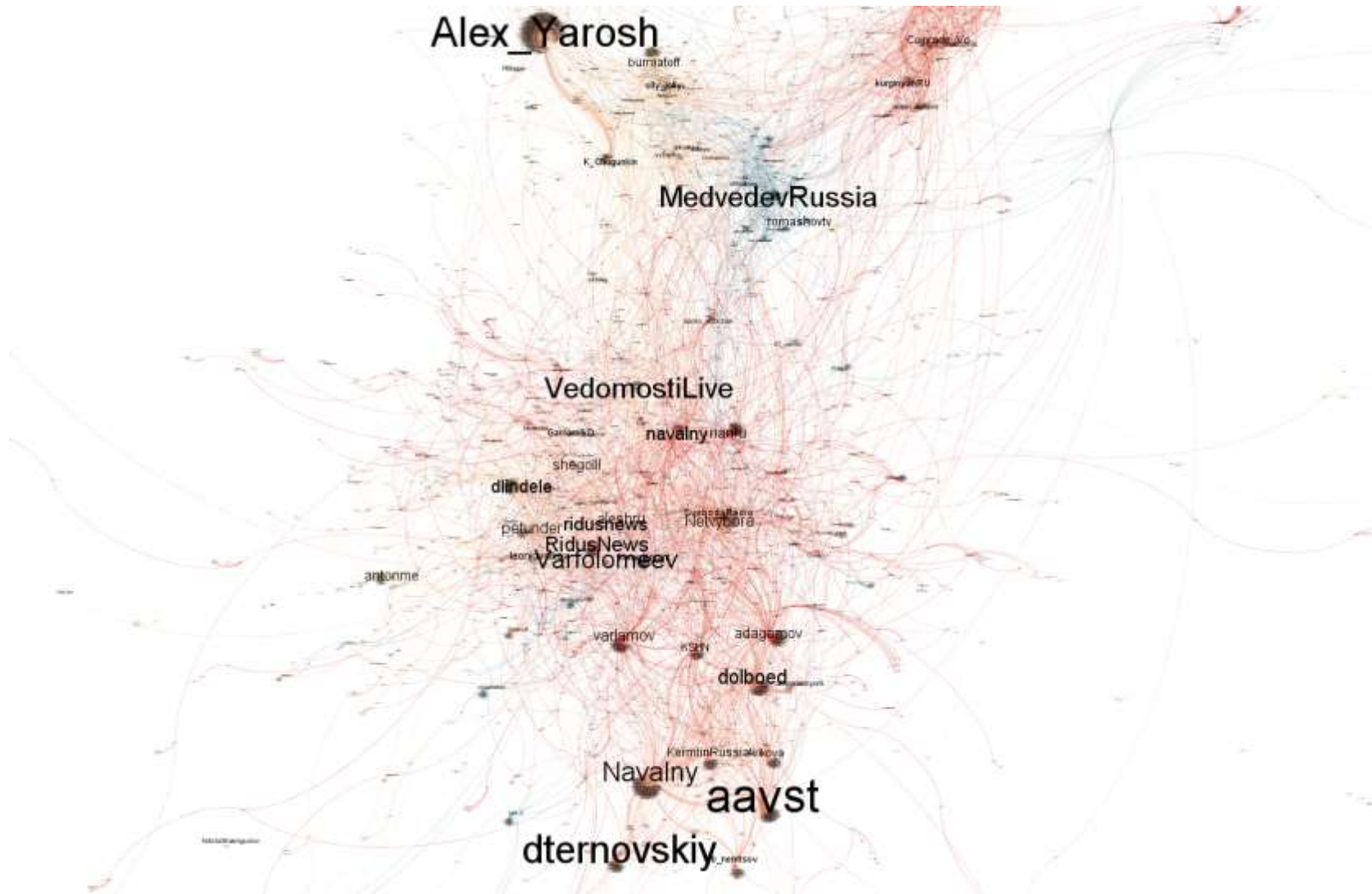
12.23.11 07:59 - 12.25.11 00:30
Timezone = UTC

	# of Messages	# of Differences	Start Date	Start Time	End Date	End Time	Replies	Retweets	Mentions
DataSift	17209	3077	23/12/2011	07:59:00	25/11/2011	00:30:00	323	923	1779

Streaming									
API	21818	6027	24/12/2011	00:00:00	24/12/2011	23:56:00	1566	5573	6508

	System Retweet			Non-system Retweet			System Reply			Non-system Reply			Retweet			Reply		
	Merged	API	Firehose	Merged	API	Firehose	Merged	API	Firehose	Merged	API	Firehose	Merged	API	Firehose	Merged	API	Firehose
Total Nodes	3267	3267	37	472	454	340	1950	1713	1939	50	44	50	3485	3469	363	1978	1739	1967
Total Edges	5576	5576	118	953	932	720	1682	1428	1673	27	23	27	6529	6508	838	1709	1451	1700
Maximum In-Degree	807	807	55	244	243	191	52	50	49	2	2	2	807	807	246	52	50	49
Maximum Out-Degree	119	119	48	58	58	59	35	35	35	2	2	2	119	119	59	35	35	35
Density	0,00	0,00	0,09	0,00	0,00	0,01	0,00	0,00	0,00	0,01	0,01	0,01	0,00	0,00	0,01	0,00	0,00	0,00
Transitivity	0,00	0,00	0,27	0,06	0,06	0,04	0,01	0,01	0,01	NaN	NaN	NaN	0,01	0,00	0,08	0,01	0,01	0,01
Diameter	9,00	9,00	3	4,00	4,00	4	7,00	7,00	7	1,00	1,00	1	13,00	14,00	4	7,00	7,00	7
Reciprocity	0,01	0,01	0,14	0,02	0,02	0,00	0,02	0,02	0,02	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,02	0,02
Degree Correlation	-0,13	-0,13	0,43	0,58	0,58	0,55	-0,03	-0,02	-0,03	1,00	1,00	1,00	-0,11	-0,11	0,55	-0,03	-0,02	-0,03

Retweet Network: how to deal with it?



3,485 unique users

24,378 tweets

12,725
mentions

6,529
retweets

Previous research

Conover et al. (2011)

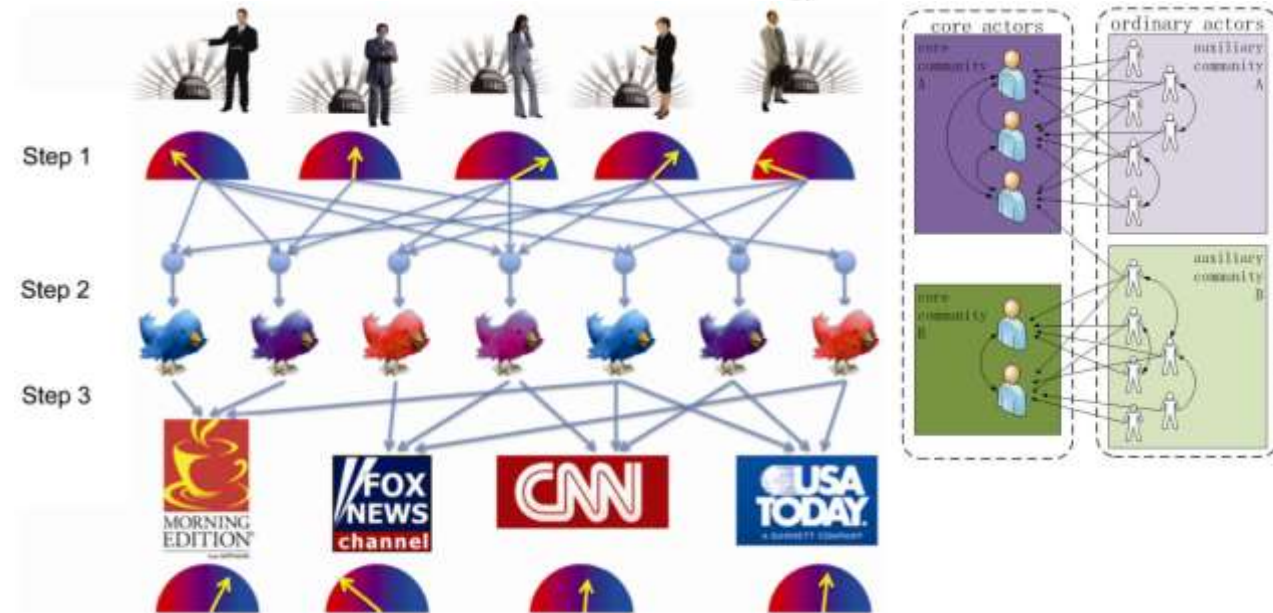
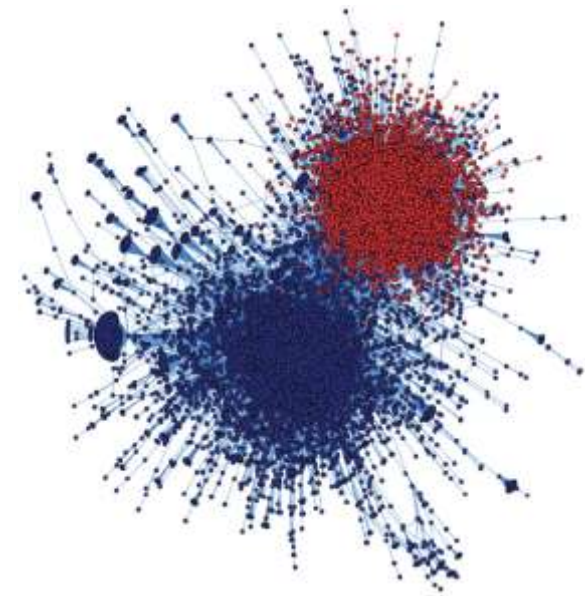
- six weeks prior to the 2010 U.S. midterm elections
- the tweets were collected by hashtags (like #p2, #tcot)
- retweet network shown rather high modularity (0.48)

Golbeck and Hansen (2014)

- dataset collected during the 111th Congress
- used known liberal / conservative ratings of Representatives
- assigned “P-scores” (political bias scores) to users based on representatives they follow

G Xun, Y Yang, L Wang, W Liu (2012)

- Some math stuff...

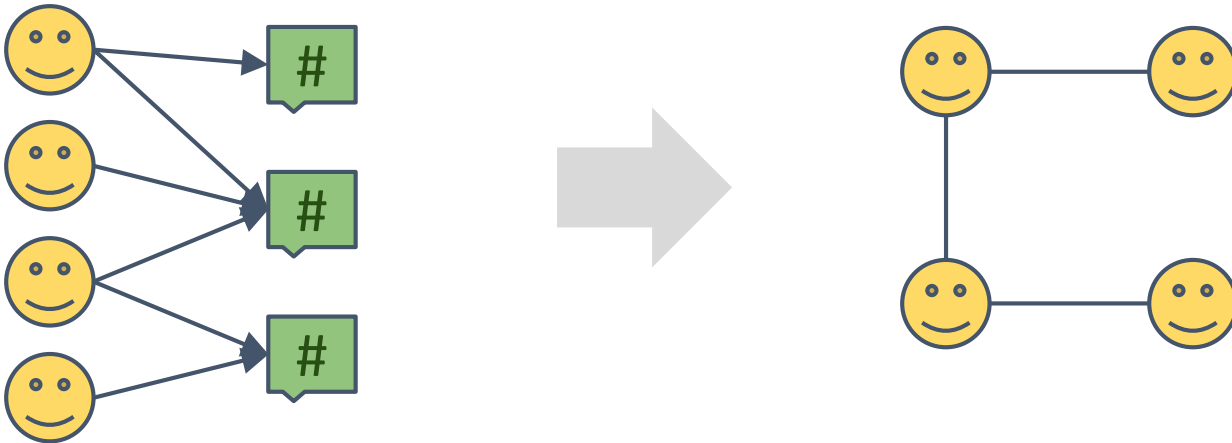


Why is our case special?

- We don't know the number of clusters (presumably more than 2);
- We don't have "follow" relationship;
- We can't reliably estimate political preferences of users, except for the most popular ones;
- There were very few political opinion-specific hashtags (and they were difficult to catch)

Bimodal networks & projections

We used natural two-mode “user-**hashtag**” and “user-**URL**” networks.
Then we projected them on the set of users using Newman (2001) method.

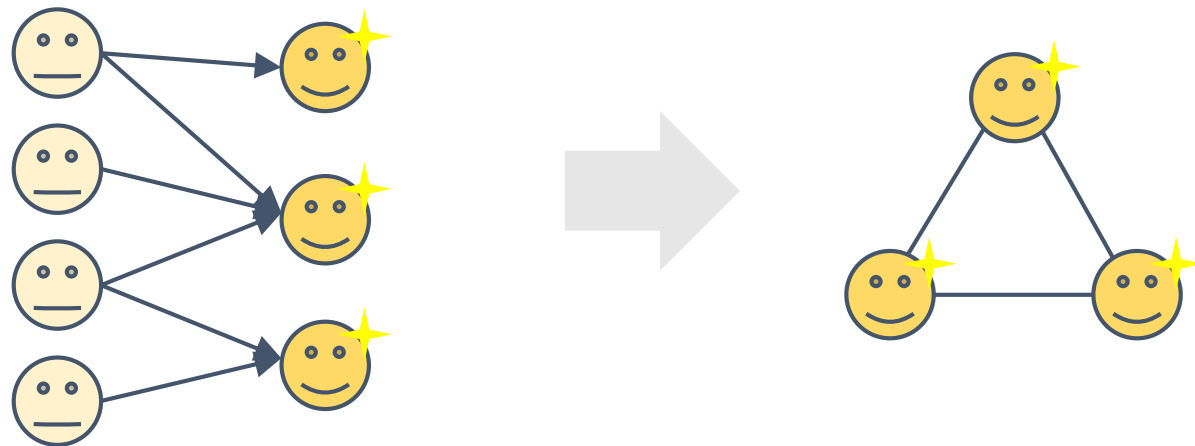


However, resulting one-mode networks had low modularity:

	Modularity
Hashtags	0.122
Links	0.485

Pseudo-two-mode networks

Top users are fairly different from the ordinary ones, so why not make them a second mode in our network? Hence, we get a **pseudo two-mode network**.



Then, we can use the same projection method to get a one-mode network of top users!

Theoretical Assumptions

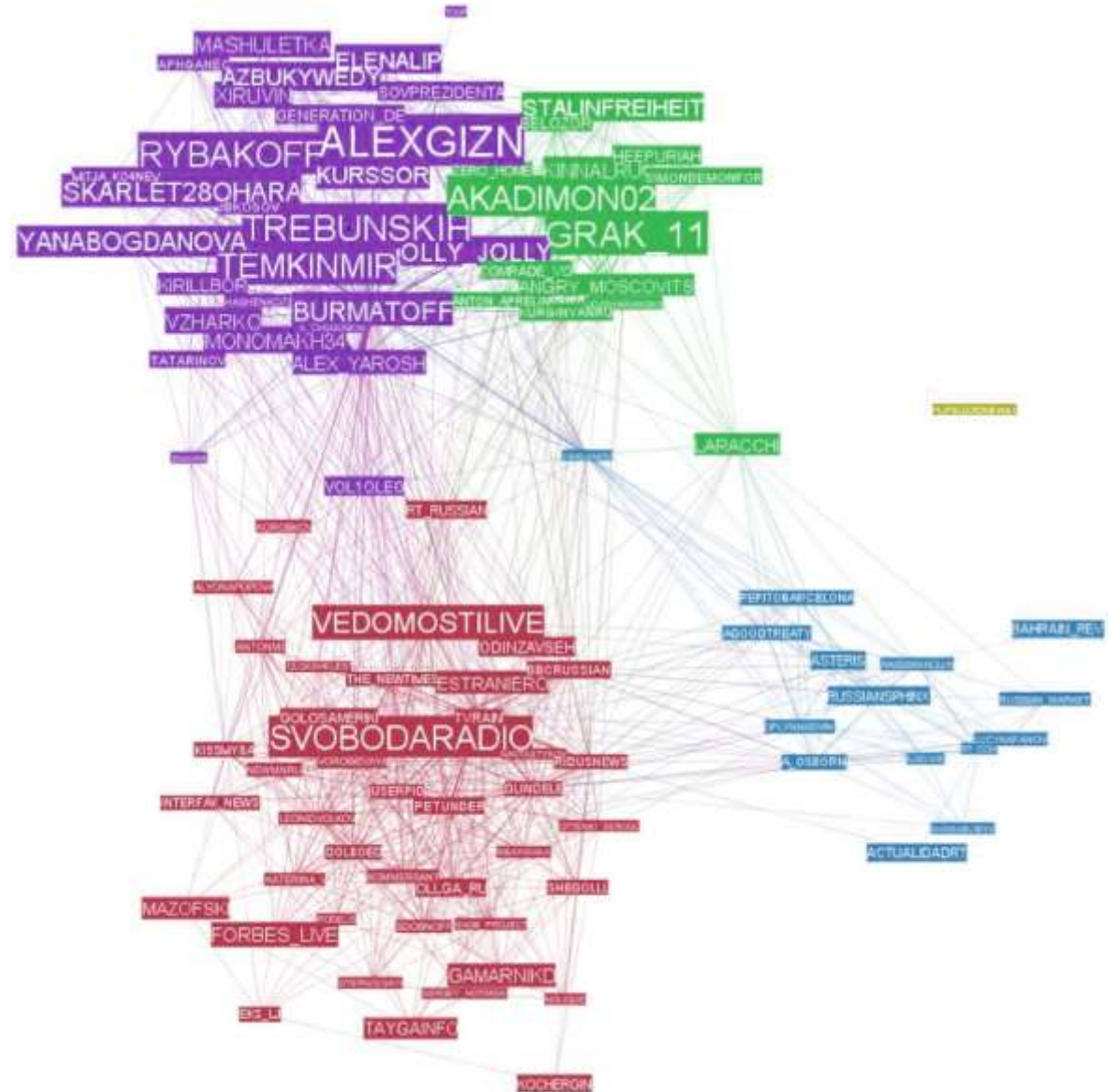
- there are users in the network with large centrality measures (see Power Law);
- they rarely interact with each other and with ordinary users;
- ordinary users tend to mention mostly top users with whose opinions they agree;
- users from both sets predominantly belong to one group each according to their opinions.

Projection results

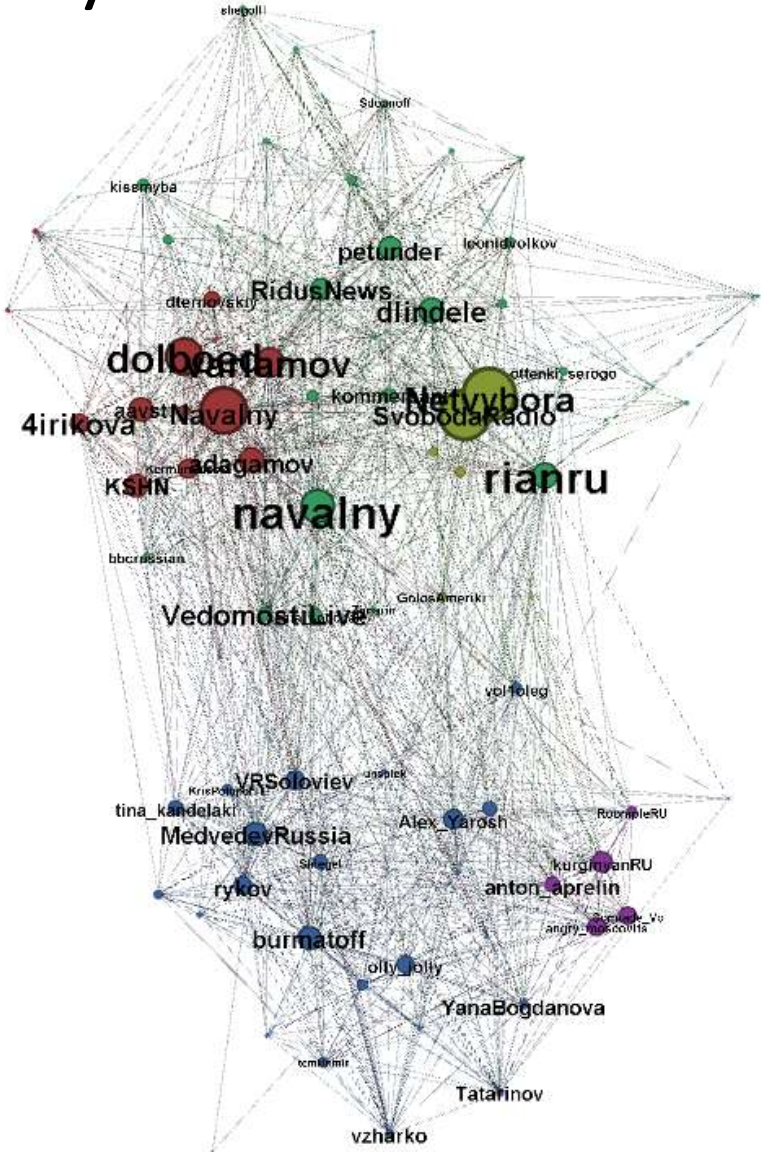
The resulting one-mode network:

- has high modularity value
- clusters (obtained with Louvain method) are clearly separable
- clusters are interpretable

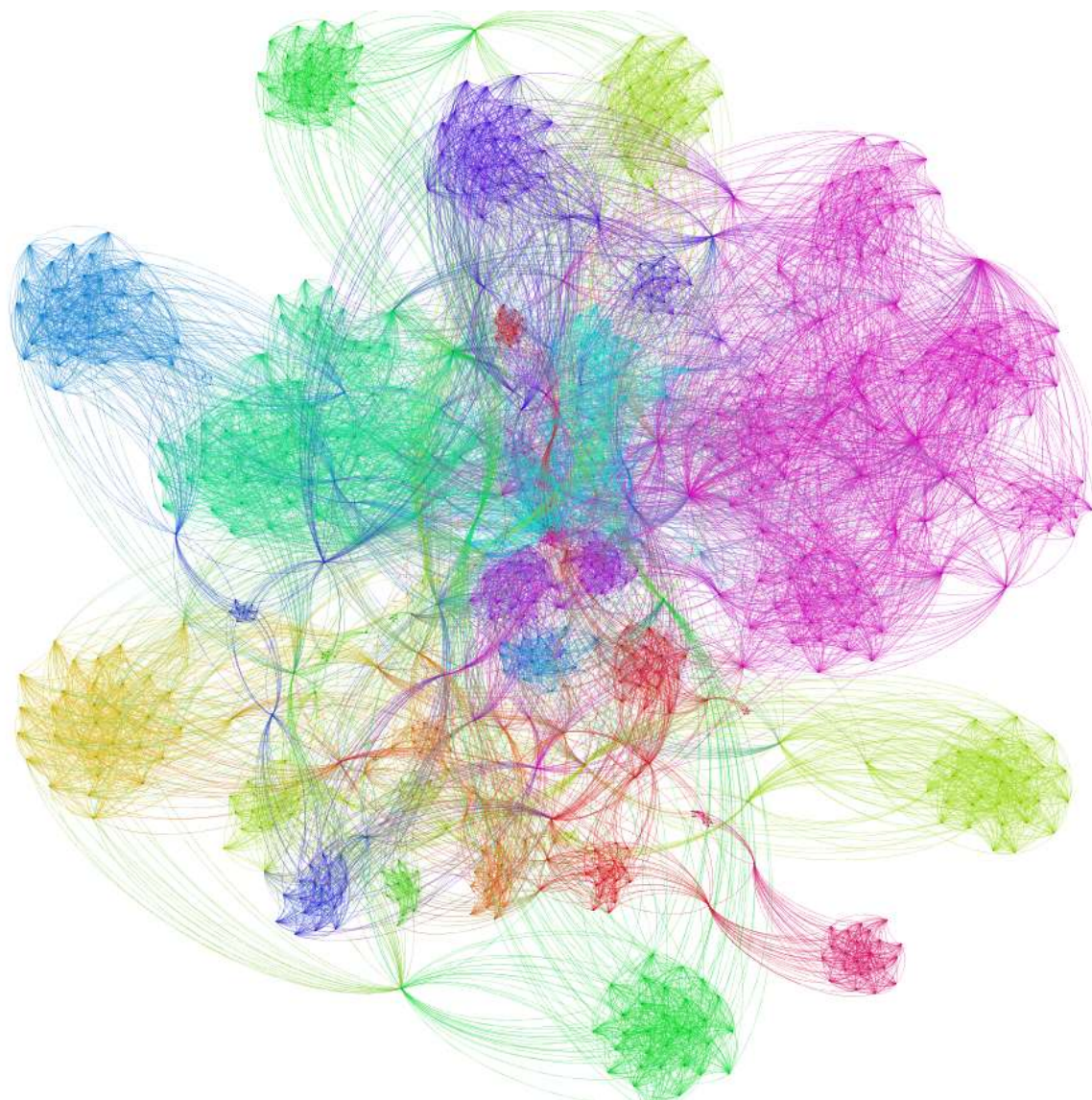
	Modularity
Mentions	0.658
Reply	0.634
Retweet	0.613



Top Users: Pretty Reasonable!



Bottom Users: How to deal with this Mess!



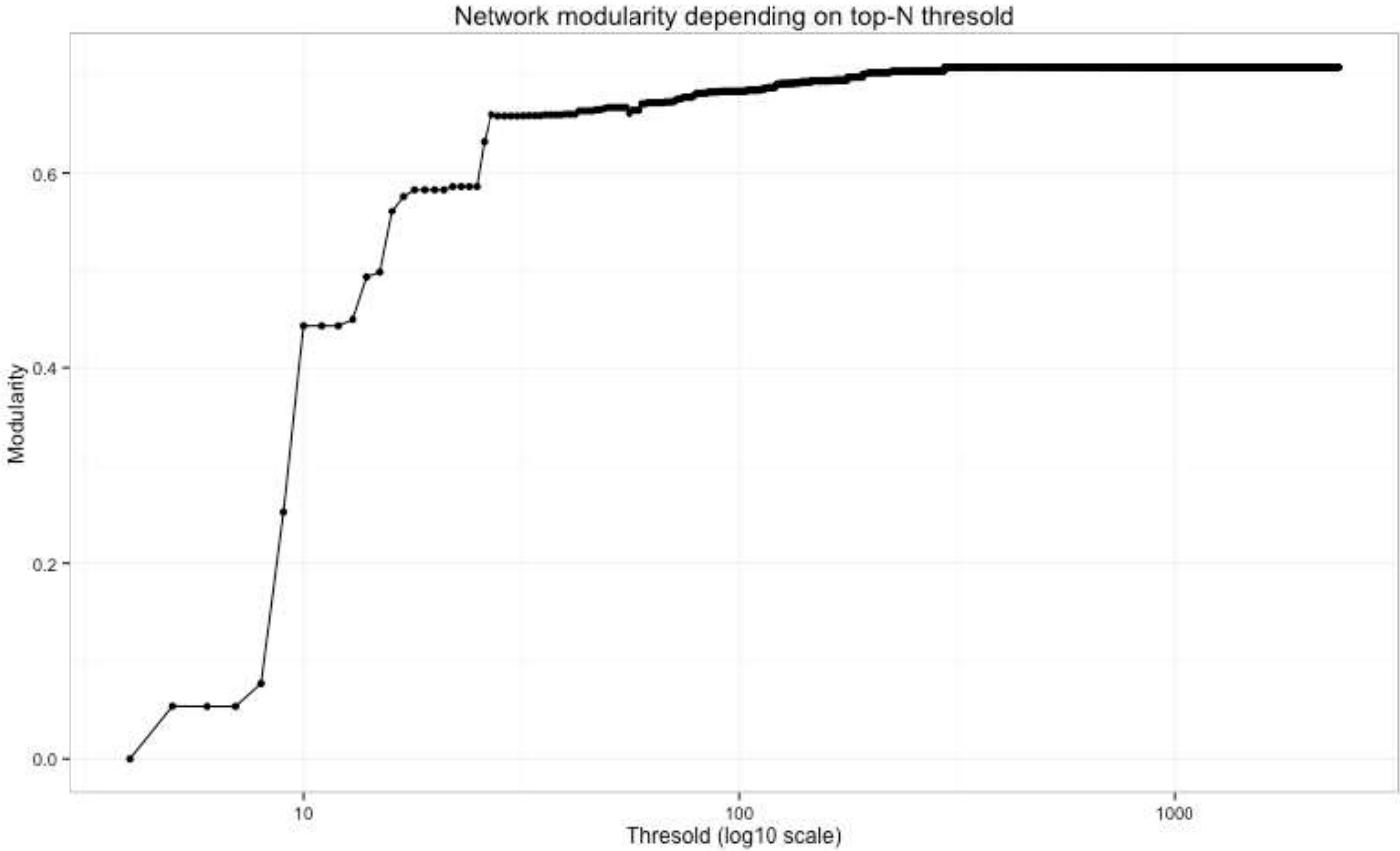
Algorithm

1. Select a set of top users for some threshold $k=1$;
 1. Make the network bimodal;
 2. Project the pseudo-bimodal network onto one of its node sets with Newman's algorithm;
 3. Perform Louvain community detection algorithm on the resulting one-mode network to calculate clusters and modularity
2. Set the threshold $k= k+1$ and repeat the steps
3. Find the optimal point where the minimal number of top users gives the best quality of clusters of the bottom users
4. Label these clusters of bottom users with the political position as the top users they were derived from

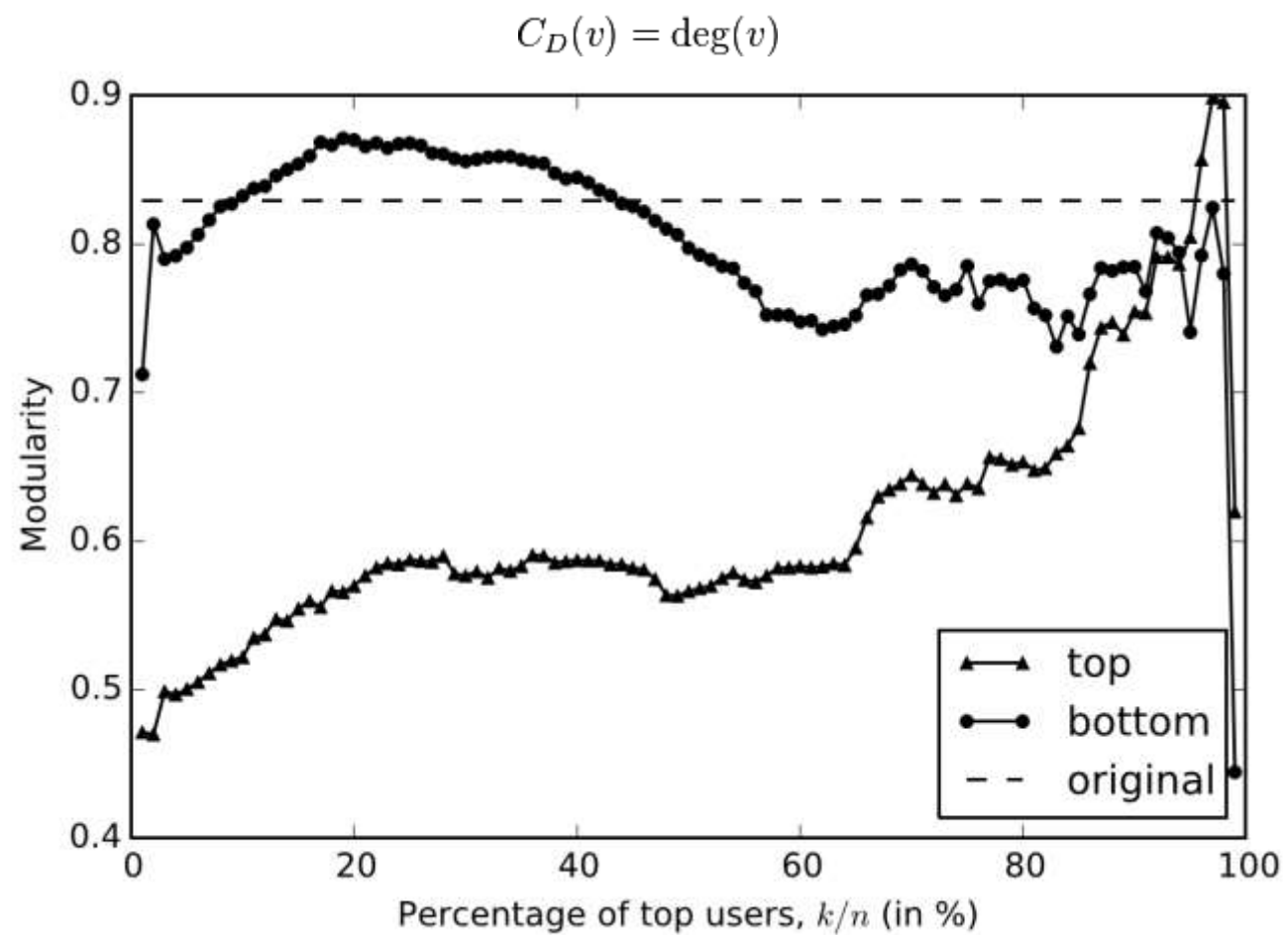
The question: which centrality metrics to choose?

- In-Degree
- Closeness
- Betweenness
- Eigenvector
- Katz centrality
- Eigenvector
- PageRank
- HITS Authority
- Bonacich Alpha
- Power Centrality

The threshold: how much is enough?

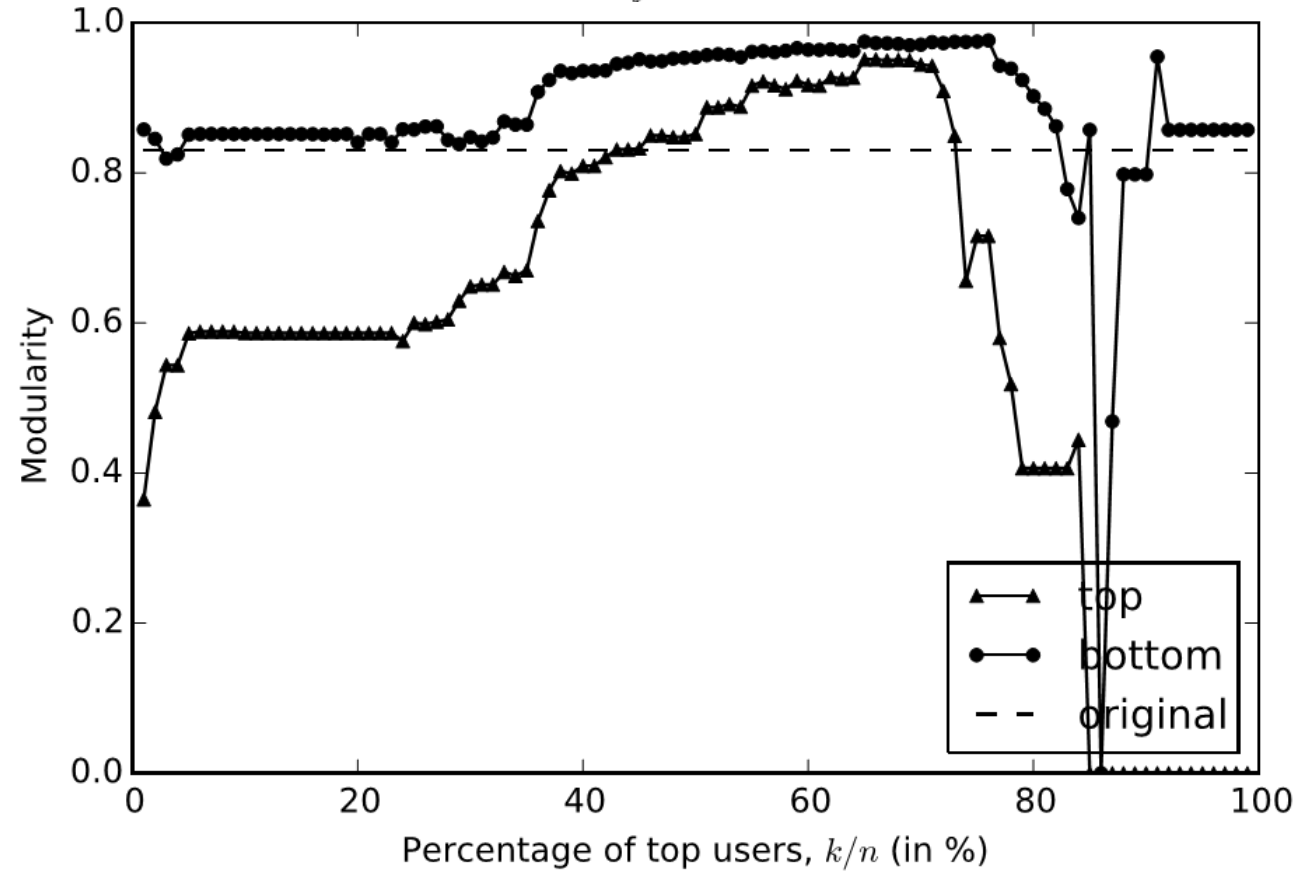


In-Degree



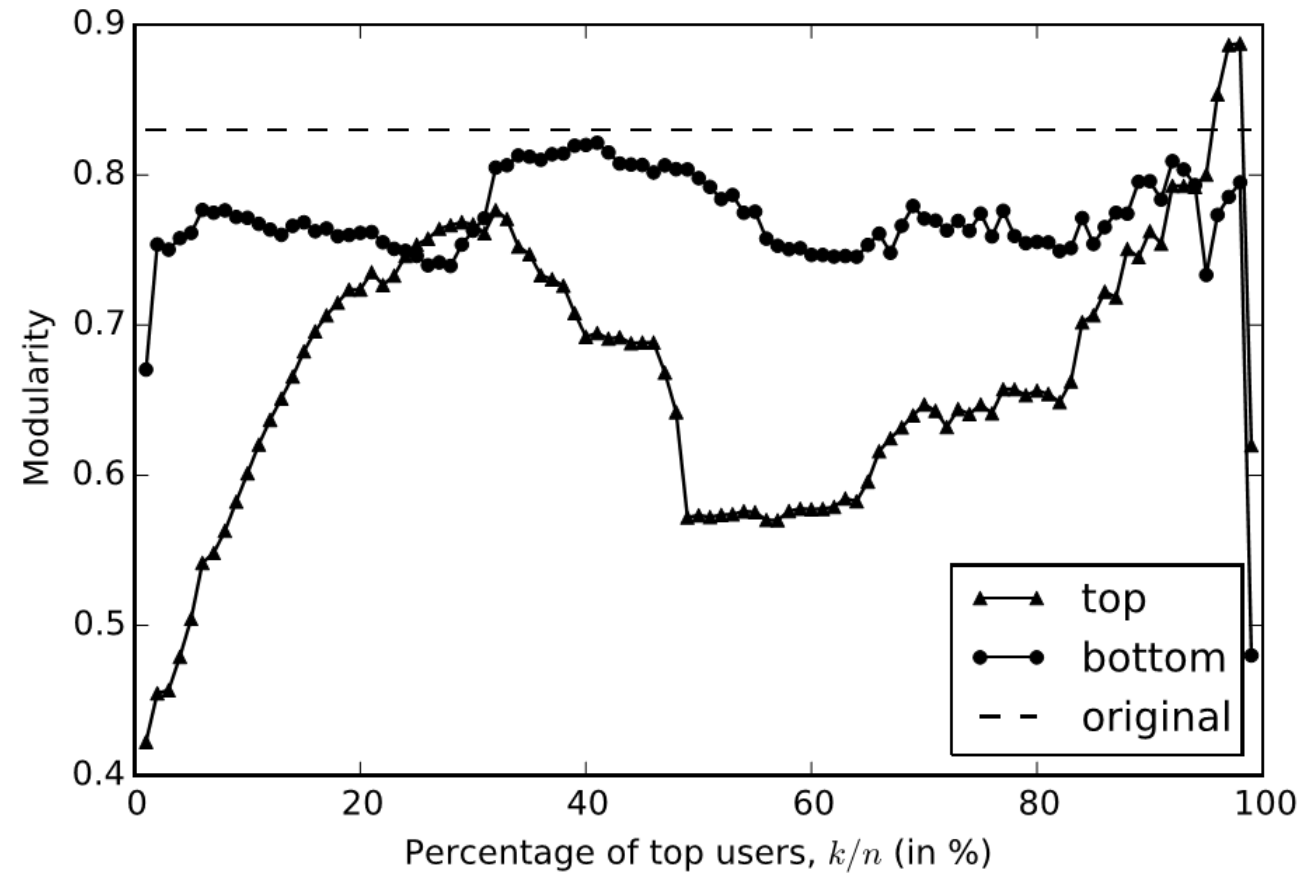
Closeness

$$C(x) = \frac{1}{\sum_y d(y, x)}$$



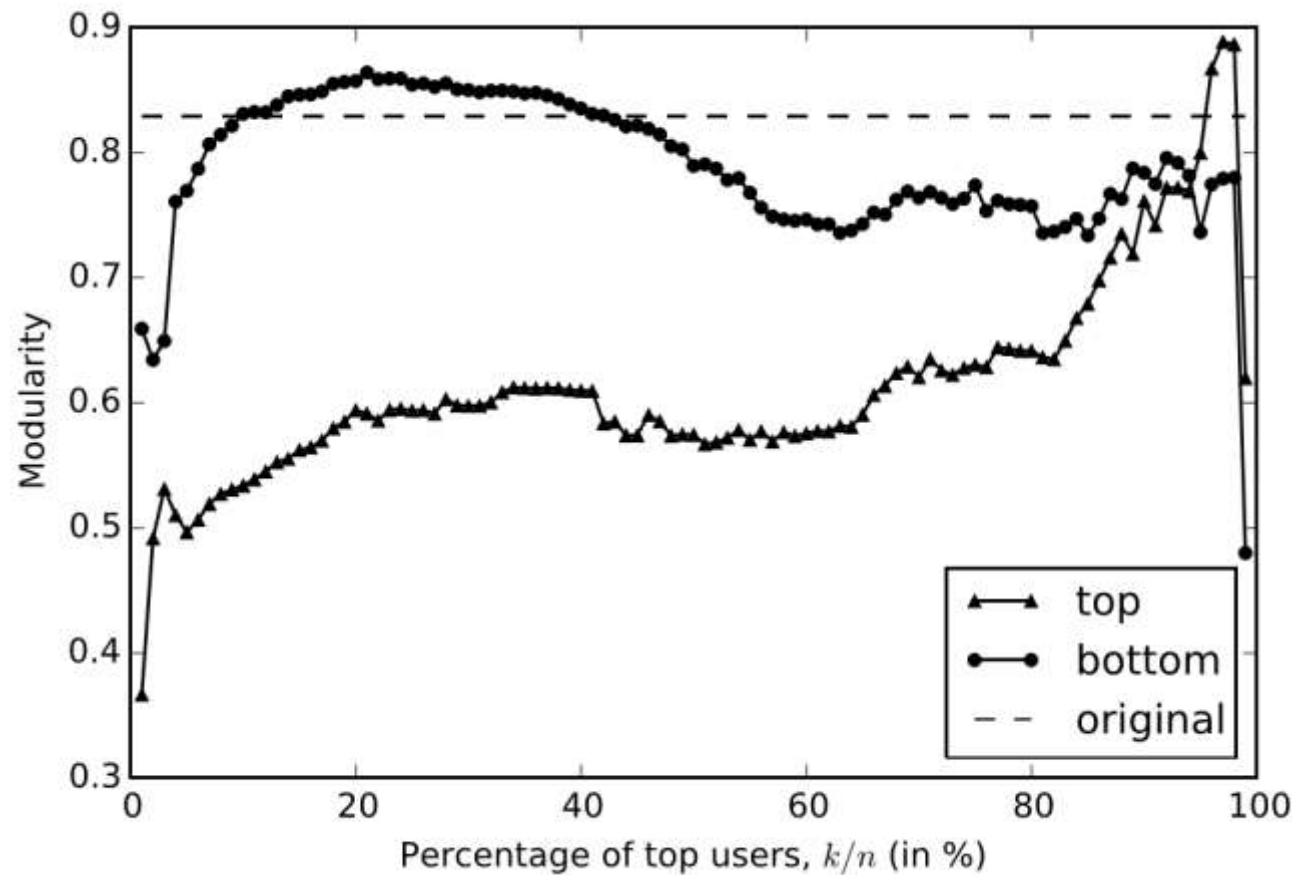
Betweenness

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



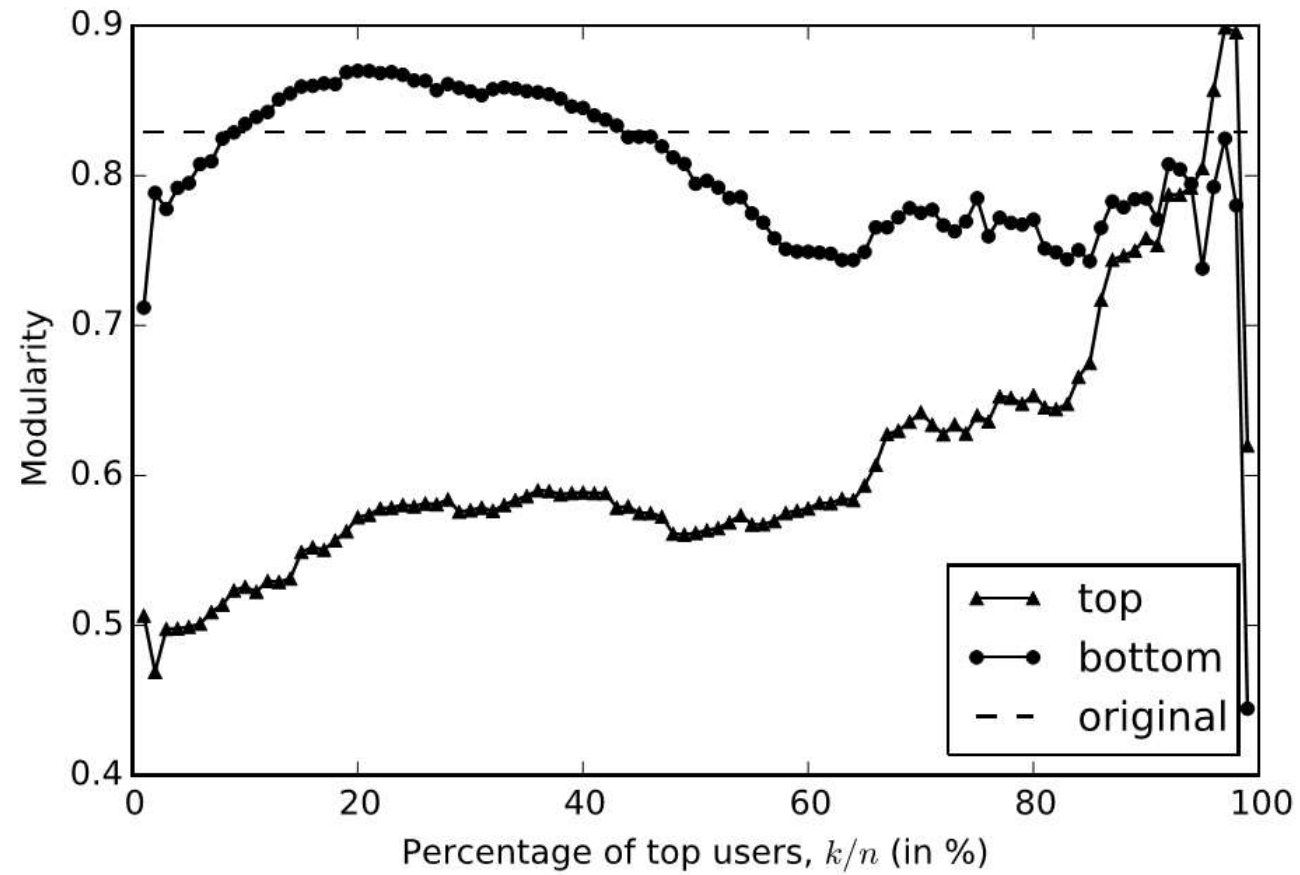
Eigenvector

$$x_i = \alpha \sum_j a_{ji} \frac{x_j}{L(j)} + \frac{1 - \alpha}{N},$$



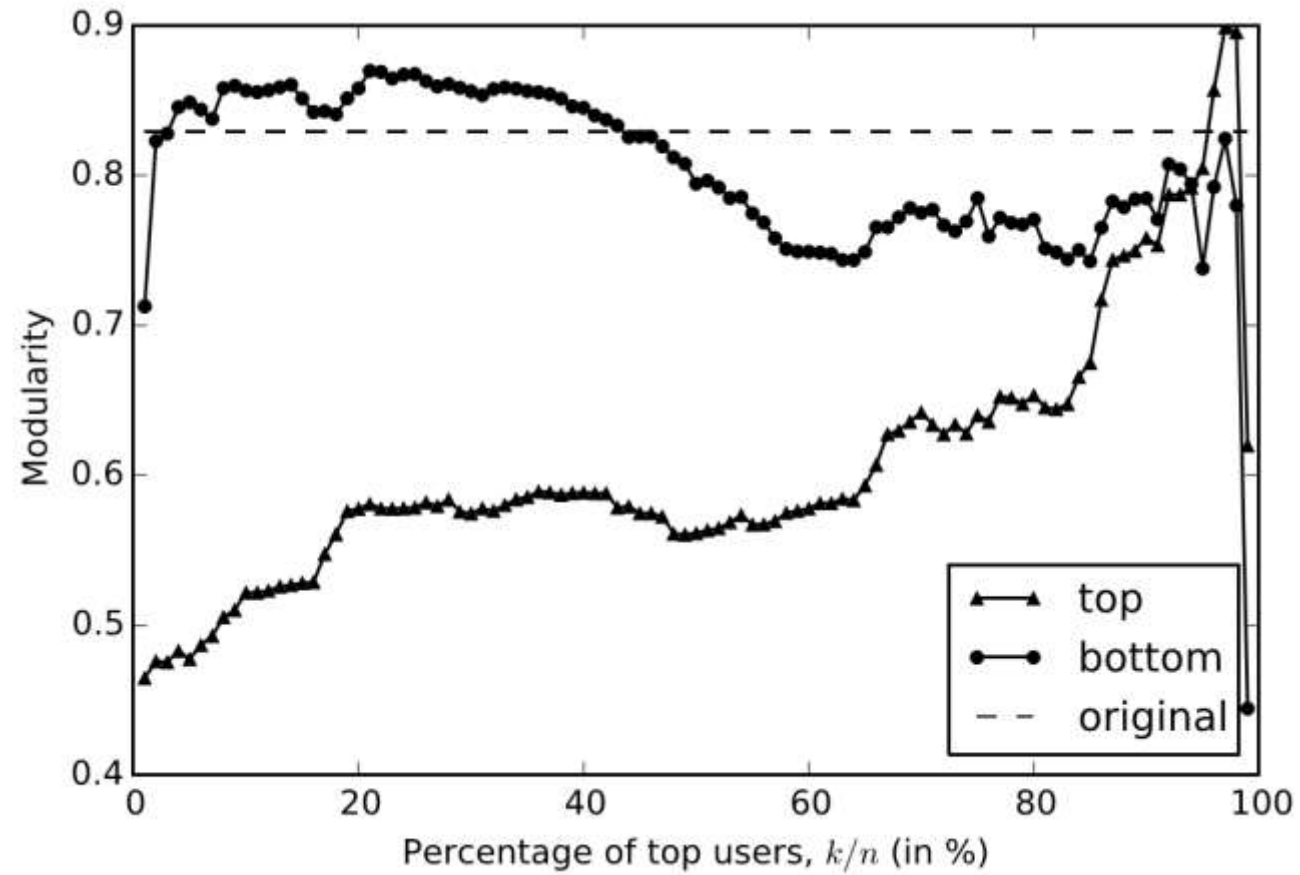
Katz Centrality

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$



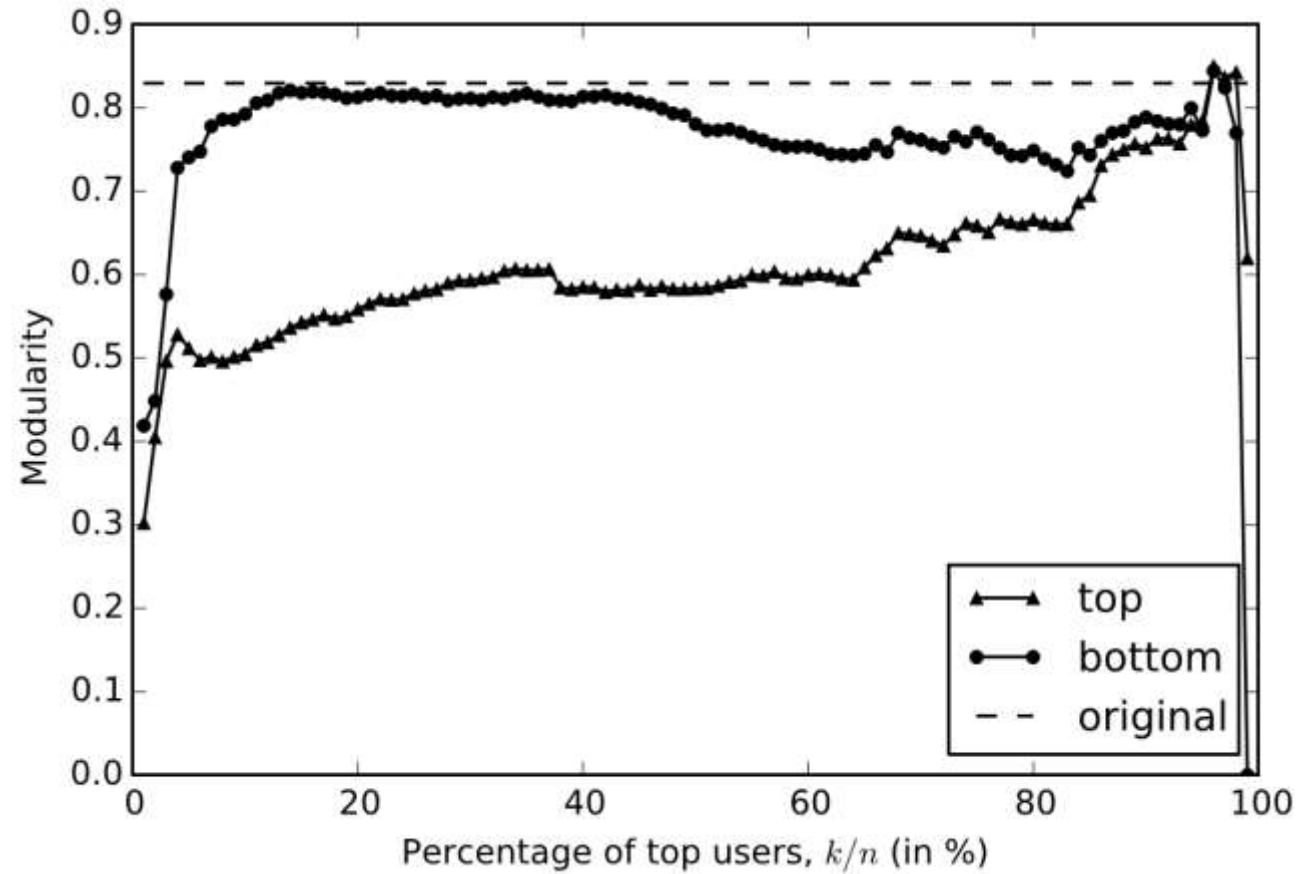
PageRank

$$x_i = \alpha \sum_j a_{ji} \frac{x_j}{L(j)} + \frac{1 - \alpha}{N},$$



HITS: Authorities

$$\text{auth}(p) = \sum_{i=1}^n \text{hub}(i)$$

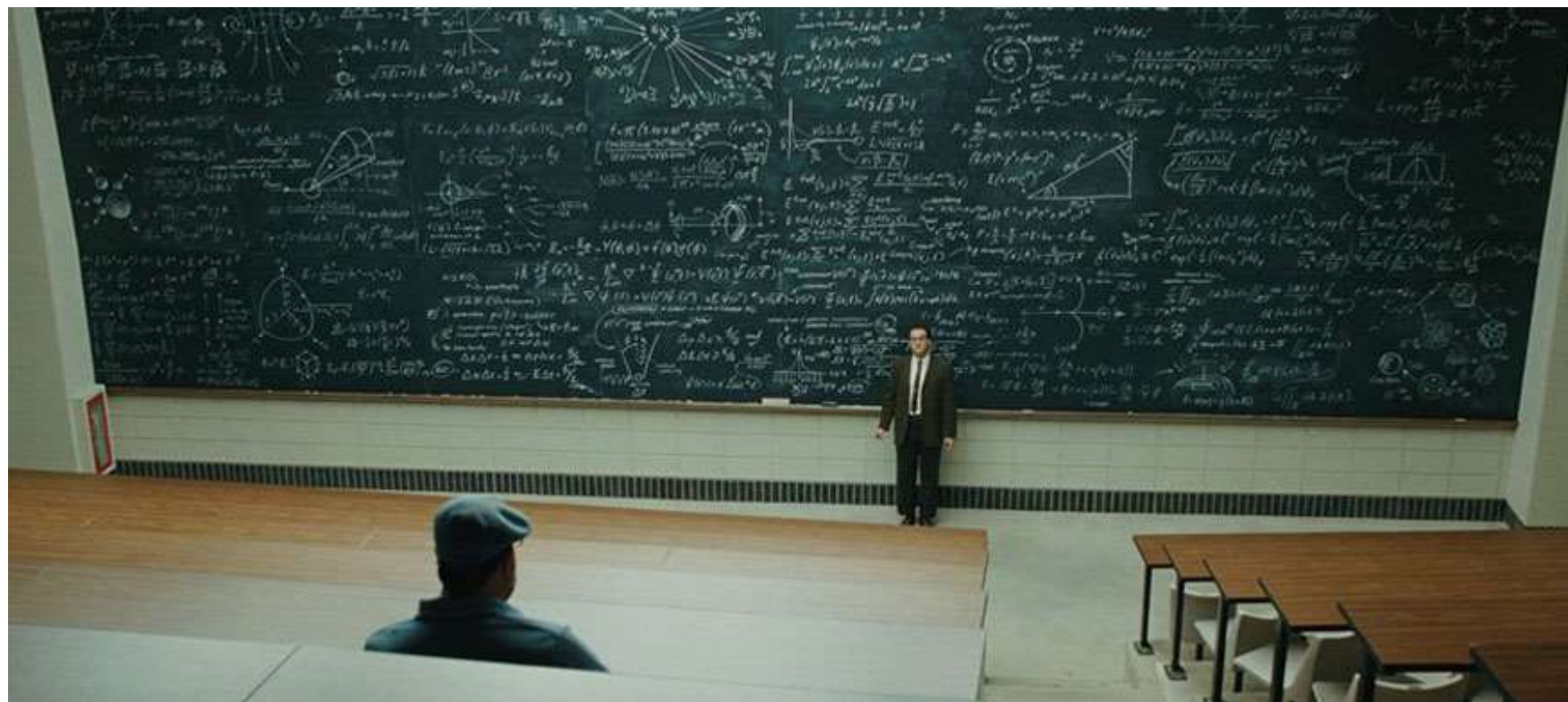


Future work (?)



- decide whether to call our network pseudo-two-mode or pseudo-bimodal (or something else)
- experiment with different strategies for selecting the set of top users (in-degree, eigenvector centrality etc.)
- analyze some similar and non-political datasets to see if distinction in modularity still holds
- Perform substantive expertise of the resulted clusters of bottom users
- Perform some text clustering techniques (LDA etc.)
- Do manual coding (if it's worth it)
- Check the accuracy of the results

Thank you for your time!



<http://jarens.ru/>

<https://twitter.com/jarens>

<http://www.pinterest.com/semenovsna/>

<http://www.scoop.it/t/social-network-analysis-by-alexander-semenov>

<ru.linkedin.com/in/semenoffalex/>