

# LABORATORY FOR INTERNET STUDIES



[www.linis.ru](http://www.linis.ru)

# PEOPLE

**Olessia  
Koltsova,**  
Head of  
laboratory



**Sergey  
Nikolenko,**  
Mathematician,  
Senior  
researcher



**Sergei Koltcov,**  
Physicist,  
technical  
director



**Peter Meylakhs,**  
Sociologist,  
Senior  
researcher



**Anastasia  
Shimorina,**  
Computer linguist,  
Junior researcher

**Svetlana  
Bodrunova,**  
Media scholar,  
Invited  
researcher



**Andrey Scherbak,**  
Political scientist,  
Invited researcher



**Yury Rykov,**  
PhD student in  
sociology,  
intern



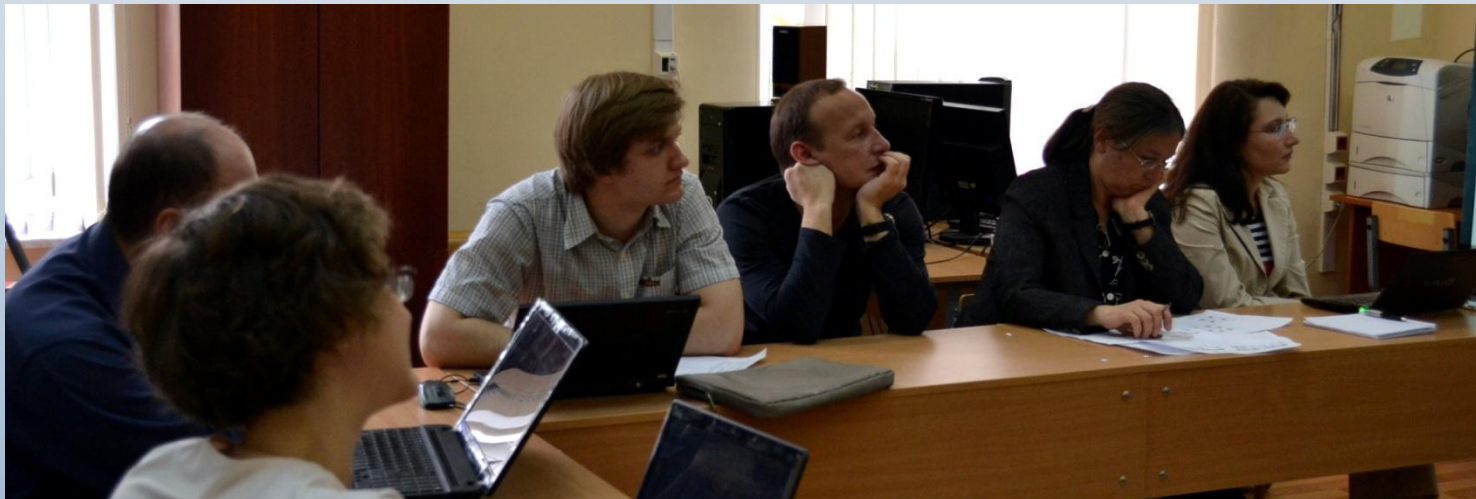
**Tatiana Yefimova,**  
PhD student in  
political science,  
intern



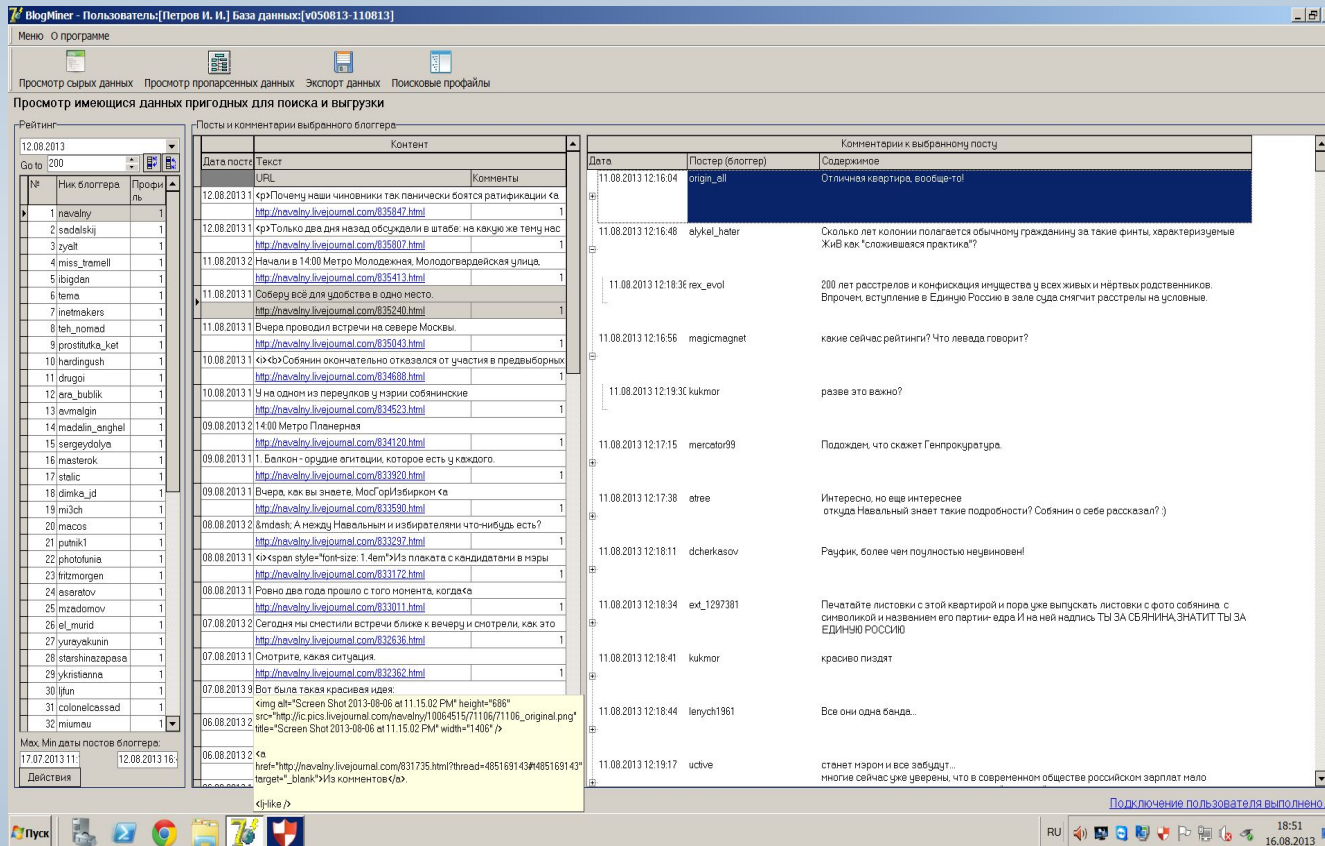
**Daria  
Yudenkova**  
Administrative  
director

# LAB'S THREE CORNERSTONES

- Big Internet data
- Diverse methods & tools
- Social science research questions and goals



# DATA1: LIVEJOURNAL



The screenshot displays the BlogMiner application window. The title bar indicates the user is 'Пользователь:[Петров И. И.]' and the database is 'База данных:[v050813-110813]'. The interface is divided into several sections:

- Top Bar:** Contains menu options like 'Меню', 'О программе', and buttons for 'Просмотр сырых данных', 'Просмотр пропарсанных данных', 'Экспорт данных', and 'Поисковые профили'.
- Left Panel:** A sidebar with a 'Рейтинг' (Rating) section showing a list of bloggers and their post counts. Below this is a 'Посты и комментарии выбранного блоггера' (Posts and comments of the selected blogger) section.
- Main Content Area:** Divided into two main columns. The left column shows a list of posts with their dates, URLs, and titles. The right column shows the full text of a selected post along with its comments.
- Bottom Panel:** A status bar showing the current date and time, and a 'Подключение пользователя выполнено' (User connection completed) message.

The posts listed in the main content area include various topics related to Russian politics and society, such as 'Почему наши чиновники так панически боятся ратификации', 'Только два дня назад обсуждали в штебе, на какую же тему нас', and 'Начали в 14:00 Метро Молодежная. Молодогвардейская улица'.

Koltran BLOGMINER mines posts,  
comments and user metadata



# DATA2: VKontakte SOCIAL NETWORK

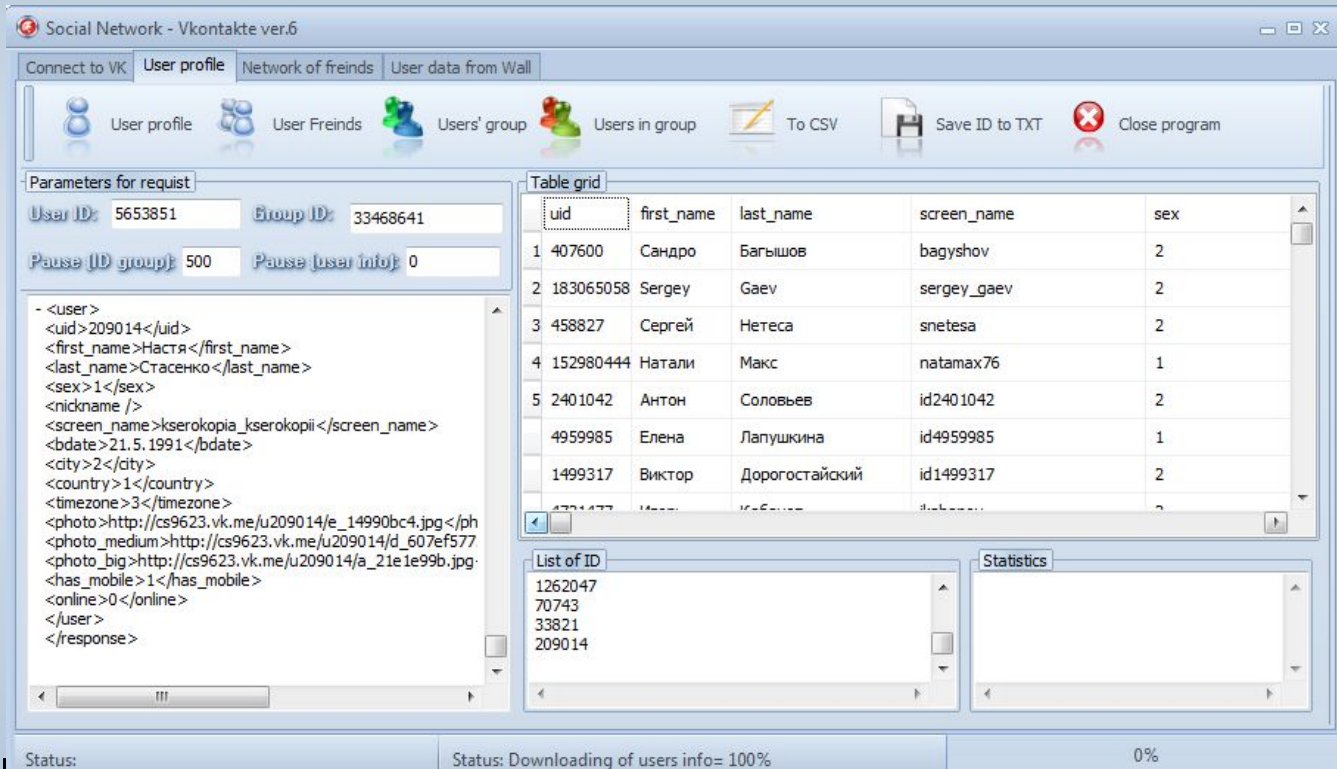


Table grid

	uid	first_name	last_name	screen_name	sex
1	407600	Сандро	Багышов	bagyshov	2
2	183065058	Sergey	Gaev	sergey_gaev	2
3	458827	Сергей	Нетеса	snetesa	2
4	152980444	Натали	Макс	natamax76	1
5	2401042	Антон	Соловьев	id2401042	2
	4959985	Елена	Лапушкина	id4959985	1
	1499317	Виктор	Дорогостайский	id1499317	2

List of ID

- 1262047
- 70743
- 33821
- 209014

Statistics

Status: Downloading of users info= 100% 0%

LINIS vkiminer mines friendship egonetworks, friendship networks of VK communities and....

Social Network - V Kontakte ver.6

Connect to VK | User profile | Network of friends | User data from Wall

DBConnection | Open Posts from DB | Read posts from InterNet | Save posts to DB | Close program

Parameters for request (getting data from wall)

Owner

☒ User ID ☐ Group ID

Read from DB

ID: 5653851

Current Post Id: 156

Posts

20.01.2013 17:39:54

23.01.2013 15:59:57 "Срочная эвакуация" отделения детской онкогематологии из 31-й Городской

24.01.2013 7:27:37

30.01.2013 5:02:17 Больница №31: чиновники зайдут с "черного хода"? : Статьи (Санкт-Петербург)

31.01.2013 4:57:53 Питерские, все на митинг в защиту 31 больницы! 3 февраля в воскресенье на

01.02.2013 13:50:20 Красная по-черному - Росбалт.py

06.02.2013 14:36:20 Друзья, у нас большое горе - умер любимый кот Барсик!(<br>на прошлой не

19.02.2013 4:51:34 Выживает не самый сильный и не самый умный, а тот, кто лучше всех приспособлен

19.02.2013 10:25:09 Вот он источник мотивации...Лучшие.

26.02.2013 6:38:27 Приятель пишет <br>Хожу в шопинг уже третьи сутки, покупаю жене платье

Comments

date	text	likes	reply_to_uid
07.02.2013	Очень жалко	0	
09.02.2013	Сочувствую	0	

Likes

post_id	owner_id	uid	copied
---------	----------	-----	--------

Status: Status 0%

...and message boards of VK communities (posts, comments, likes, user metadata).

# TOOLS & METHODS 1

## Text mining: topic modeling with LINIS TopicMiner

TopicMiner ver. 41

Stemming | Gibbs LDA sampling | Kullback Leibler Distance

Parameters of simulation

Alpha: 0,5  
Beta: 0,1  
Save step: 50

Number of documents: Label1

Perplexity

Calculation method

Words with high probability

	1	2	3	4	5	6
1	федерация: 0,0147	верба: 0,007033	мрамор: 0,015868	mdash: 0,016396	танк: 0,018219	японский: 0,0157
2	российский: 0,012	ад: 0,005830	соотечественник: 0,0	мост: 0,008951	т: 0,006409	фонд: 0,010658
3	трудоу: 0,00778	хусаиновна: 0,00	ваниль: 0,007334	остров: 0,008124	мм: 0,006021	япония: 0,00916
4	к: 0,007614	женщина: 0,0044	стручок: 0,004287	тоннель: 0,006304	вооружение: 0,00544	только: 0,008497
5	работодатель: 0,0	ангелин: 0,00422	месторождение: 0,00	остер: 0,005642	машин: 0,005247	вот: 0,008331
6	район: 0,006578	террорист: 0,0036	гражданство: 0,00408	поезд: 0,005477	пушка: 0,004666	посольство: 0,00
7	статья: 0,006233	ветка: 0,003226	камень: 0,004084	строительство: 0,0	название: 0,004472	деньга: 0,006834
8	работник: 0,00606	земля: 0,003025	разный: 0,003271	железнодорожный	башня: 0,004279	пономарев: 0,00
9	расход: 0,005715	дерево: 0,002825	нация: 0,003068	шоколад: 0,002995	польский: 0,003504	нужный: 0,00650
10	стоимость: 0,0057	почка: 0,002625	мраморный: 0,002865	км: 0,002995	армия: 0,003504	американский: 0,0
11	север: 0,005369	освящать: 0,0022	гражданин: 0,002865	бисквит: 0,002829	украинский: 0,00331	британский: 0,00
12	отпуск: 0,005369	душ: 0,002024	РВП: 0,002662	стоимость: 0,0028	снаряд: 0,003117	правда: 0,00550
13	компенсация: 0,00	веточка: 0,00182	россия: 0,002662	шоколадный: 0,002	комплекс: 0,003117	японец: 0,00517
14	конституционный:	век: 0,001823	документ: 0,002458	тут: 0,002664	орудие: 0,002924	пекин: 0,005005
15	крайний: 0,005024	вход: 0,001823	квота: 0,002255	линия: 0,002498	платформа: 0,00234	восток: 0,005005
16	договор: 0,004852	праздник: 0,0018	полчение: 0,002255	ччасток: 0,002333	таварс: 0,002149	весчср: 0,004835

Export to Excel...

Number of words for export: 100

Status: Calculation completed

# TOOLS & METHODS 2

## Text mining: sentiment analysis

Russian language  
tools still need to be  
developed

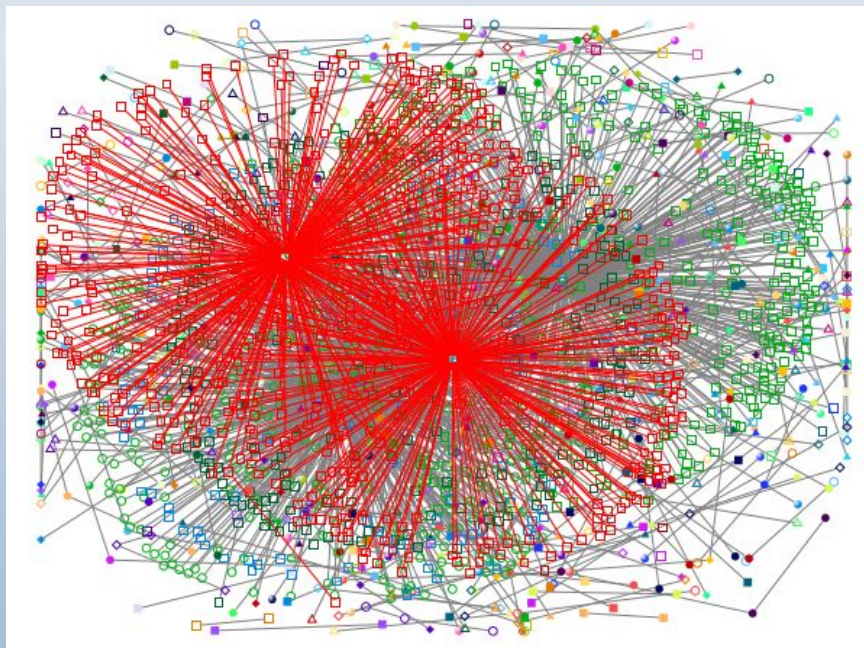
Sentiment lexicon will  
be constructed and  
tested with  
SentiStrength (Thelwall)





# TOOLS & METHODS 3

## Network analysis



Many tools exist, but most do not process data of our scale

Community detection with Louvain source code

# RESEARCH QUESTIONS AND GOALS 1

- Online public agenda: topical structure of the Russian blogosphere and influence of national elections 2011-2012 on it (finished project)



# RESEARCH QUESTIONS AND GOALS 3



**НЕАНДЕРТАЛЬЦЫ**  
теперь у тебя во дворе

demotivation.ru

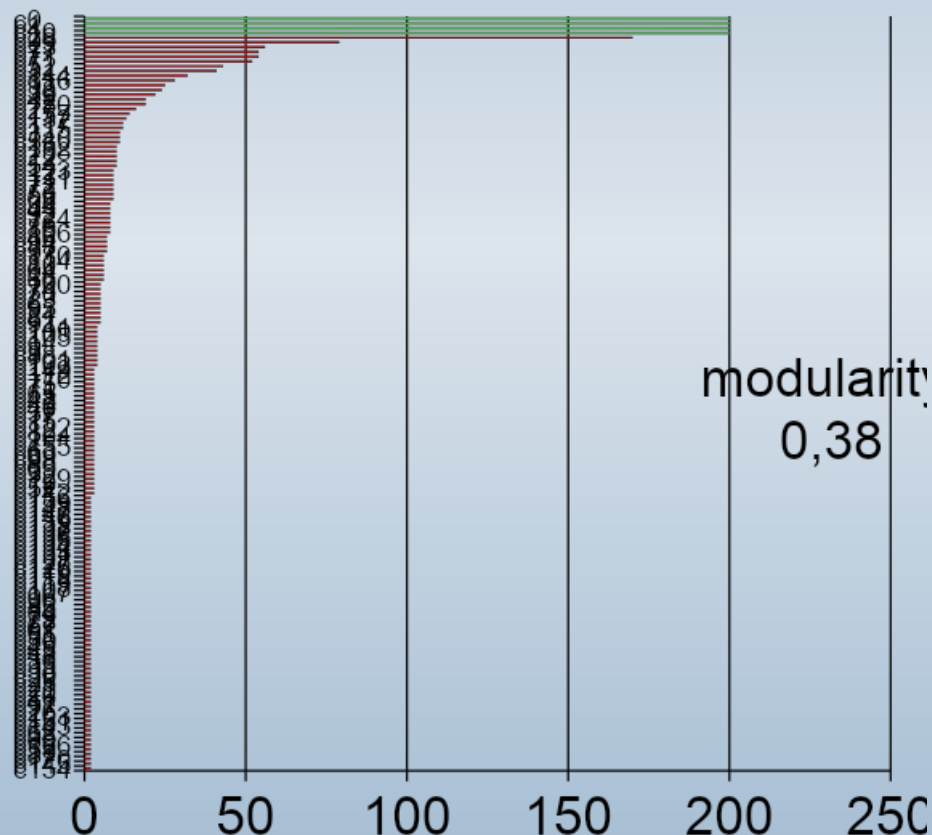
“Neanderthals are now in your backyard” (a picture from a blog post concerning Muslims slaughtering sheep for Kurban-Bayram in Moscow)

Online public opinion:  
revealing topics that  
arouse social tension  
through sentiment  
analysis

(pilot project finished;  
main project planned)

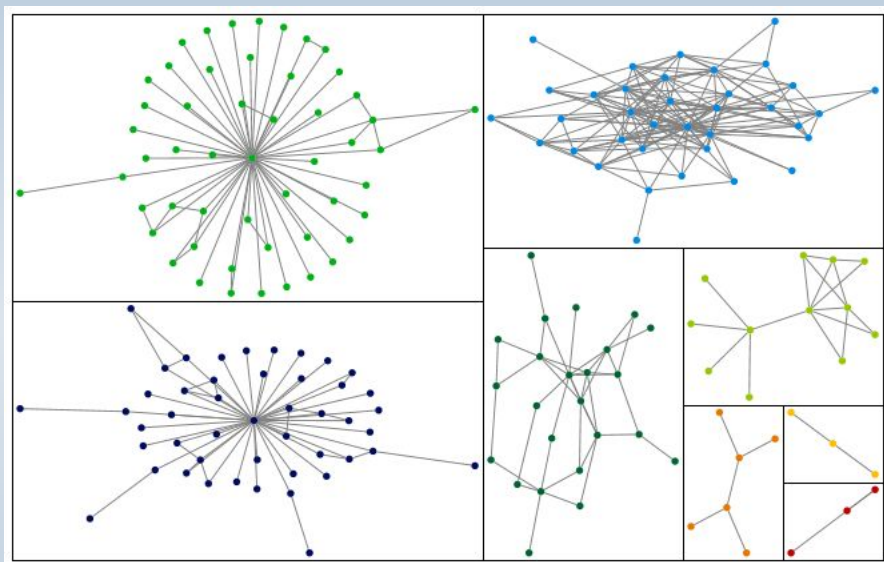
# RESEARCH QUESTIONS AND GOALS 3

Comment-based  
communities,  
their topical  
coherence and  
authorship  
homogeneity  
(finished project)



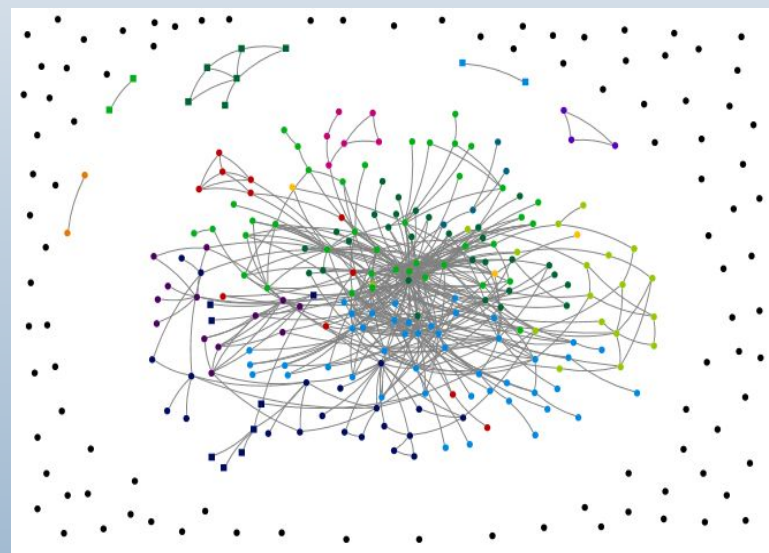


# RESEARCH QUESTIONS AND GOALS 4



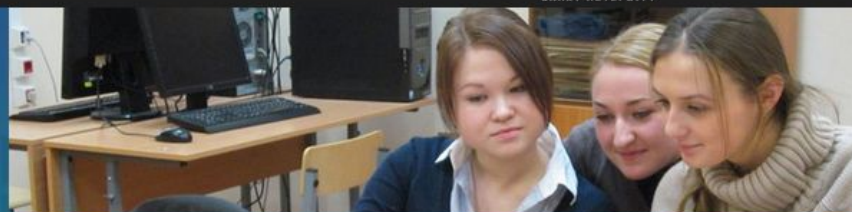
Network structure  
and leadership in VK  
communities of  
social movements

Influentials and  
suscep-tibles in  
destructive com-munities:  
perspectives for Internet



# THANK YOU!

← → ↺ [www.linis.hse.spb.ru/index.php/glavnaja.html](http://www.linis.hse.spb.ru/index.php/glavnaja.html)



[Главная](#)

[О нас](#)

[События](#)

[Публикации](#)

[Исследования](#)

[Материалы](#)

[Ссылки](#)

[Контакты](#)

[ENGLISH](#)

 [»](#)


## Новости

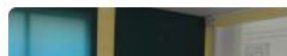


### Чем дышит блогосфера? Семинар ЛИНИС в Москве

11:54 27.04.2012

24 апреля команда ЛИНИС провела в НИУ-ВШЭ (Москва) семинар на тему "Чем дышит блогосфера? Методы анализа больших массивов Интернет-данных для социологических задач". Выступление прошло в рамках академического семинара по социологической теории и методологии кафедры анализа социальных институтов (рук. Инна Девятко), который на этот раз проходил совместно с Лабораторией экономико-социологических исследований (рук. В.Радзев).

[Подробнее...](#)



### ЛИНИС – участник Балтийского партнерства по новым медиа

12.05.2012



## Анонсы

11.05.2012 (Пятница)

### Презентация Лаборатории Интернет-Исследований

11 мая в 17-00 состоится презентация новой исследовательской площадки Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург) - Лаборатории интернет-исследований (ЛИНИС).

[Подробнее...](#)

26.09.2012 (Среда)

### Новые СМИ: меняющийся медийный ландшафт

27-28 сентября 2012 года лаборатория интернет-исследований совместно с зарубежными партнерами проводит конференцию "New media: changing media landscapes". Конференция