

«Social capital of nascent entrepreneur and its influence on their projects' success: online component»

Dornostup Olga, Suvorova Alena
St.Petersburg School of Economics and
Management

Problem

External indicators of entrepreneurial risk:

- *political and economic conditions*
- *tax system*
- *competition*
- *etc.*

Internal characteristics:

- *competence and knowledge*
- *level of entrepreneurial culture*
- *social capital*
 - offline
 - online

Research questions

1st: startupper's behavior on social networking site differs from random users' behavior

2nd: potential success of start-ups relates to a degree of entrepreneur's activity, as a user of the social network

Nascent entrepreneurs:

1672
(projects from Russian
Startup Rating)



1564
(unique startupper)



623
(were found in Vk.com)

Random users of Vk.com:



Network-related
features of **912584**
users



Content-related
features of **22580**
users

Output variables

Variable	Scale
<i>Rating</i>	from D to AAA (highest) [10 levels]
<i>Grade</i>	Low, medium, high [3 levels]
<i>Team rating</i>	from 0 to 5 (highest) in 0.5 steps [11 levels]
<i>Product rating</i>	
<i>Experts rating</i>	
<i>Finance rating</i>	
<i>Law rating</i>	
<i>PR rating</i>	

Input variables

Network-related

Number of friends

Number of groups

Number of followers



Startup-related

Startup-friends

Number of top30 groups

Content-related

All posts

Owner's posts

All comments

Owner's comments

Likes on all posts

Likes on owner's posts

Personal info

Gender

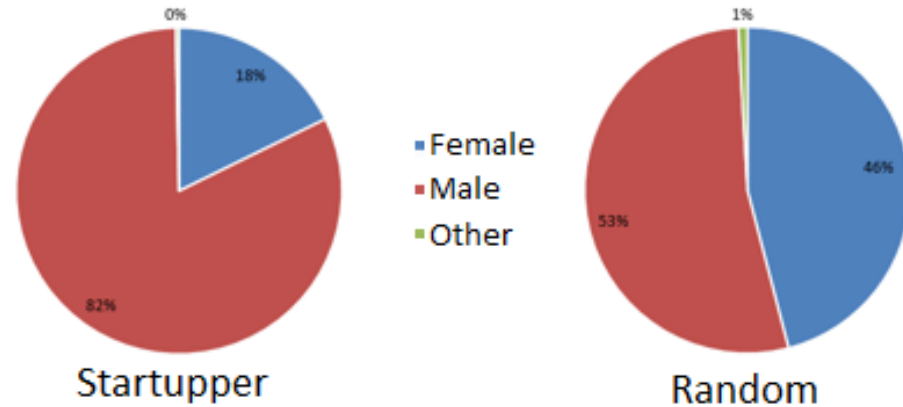
Age

City

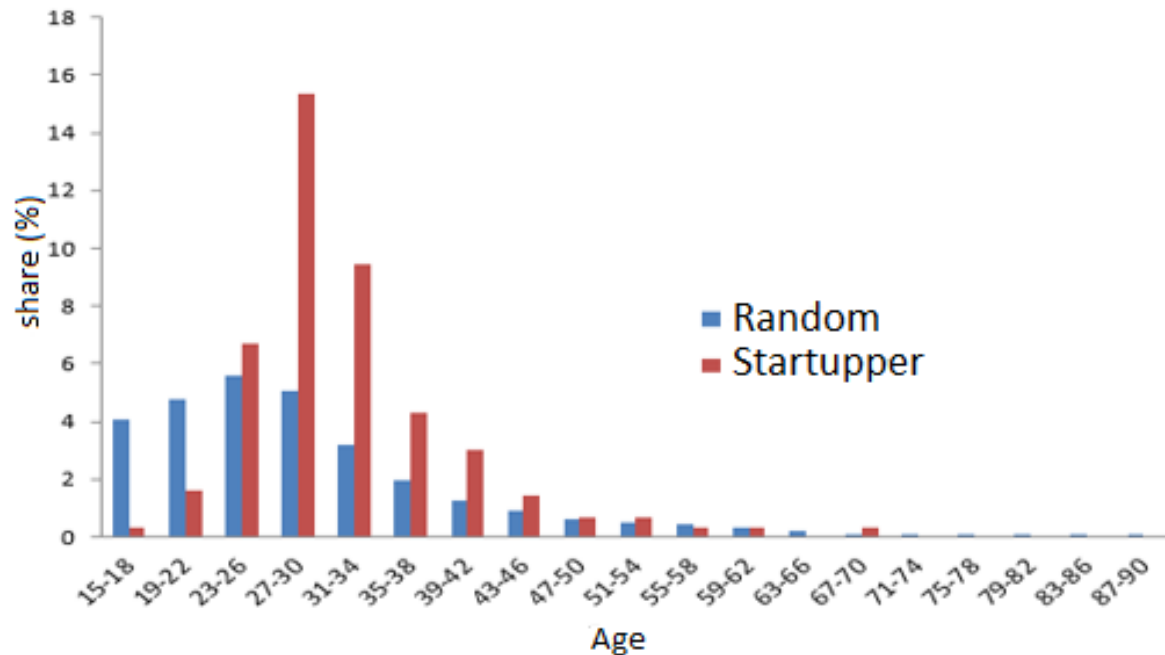
Country

Results

- Gender distribution



- Age distribution



Results

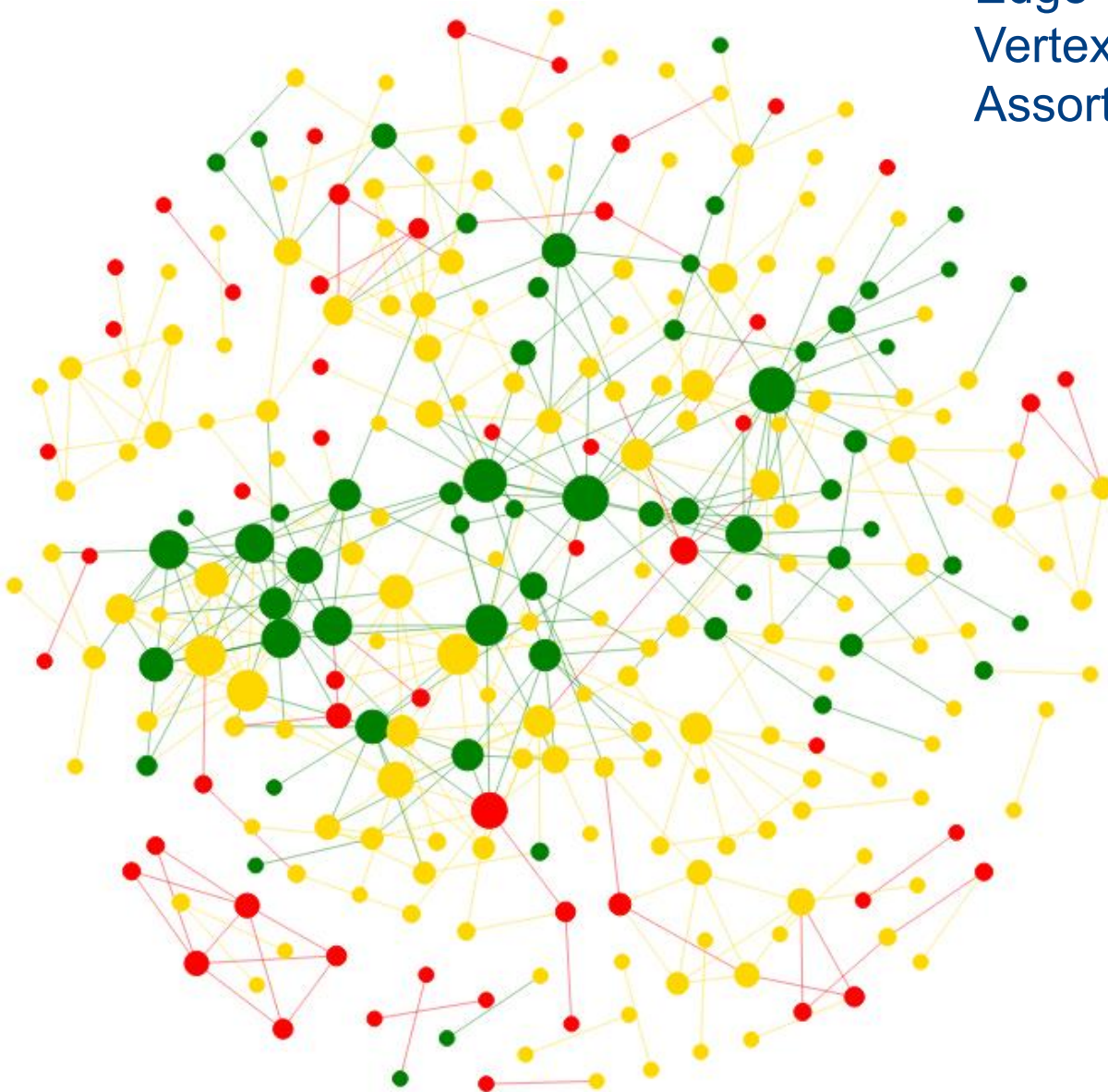
Input variable	Odds ratio
<i>Friends</i>	1.0100***
<i>Groups</i>	0.9949***
<i>Followers</i>	1.0002
<i>Other's posts</i>	1.0002***
<i>Owner's posts</i>	1.0005***
<i>Other's comments</i>	1.0003***
<i>Owner's comments</i>	1.0010***
<i>Likes on other's posts</i>	1.0002***
<i>Likes on owner's posts</i>	1.0003***

Grade distribution

Edge = 422

Vertex = 273

Assortativity (grade) = 0.0832



Grade:

- - high
- - medium
- - low

Results

Pearson's correlations

Variable	Rating	Team	Experts	PR	Finance	Law	Product
<i>Friends</i>		0,09	0,08				
<i>Followers</i>							
<i>Groups</i>							
<i>Startup-friends</i>	0.2	0,09	0,12				
<i>Top30 groups</i>	0.1	0,1	0,13			0,07	
<i>Other's comments</i>				0,09	0,12		
<i>Owner's comments</i>				0,09	0,12		
<i>Other's posts</i>				0,12	0,09		
<i>Owner's posts</i>	0.08			0,12			
<i>Likes on other's posts</i>				0,13			
<i>Likes on owner's posts</i>	0.11			0,12			

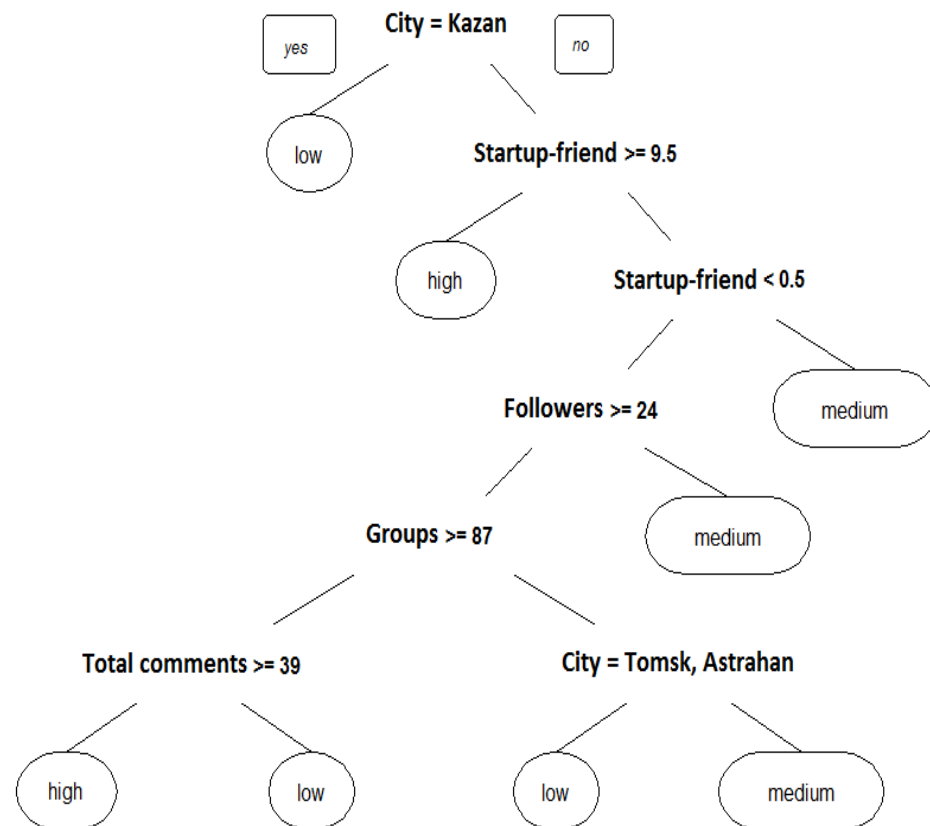
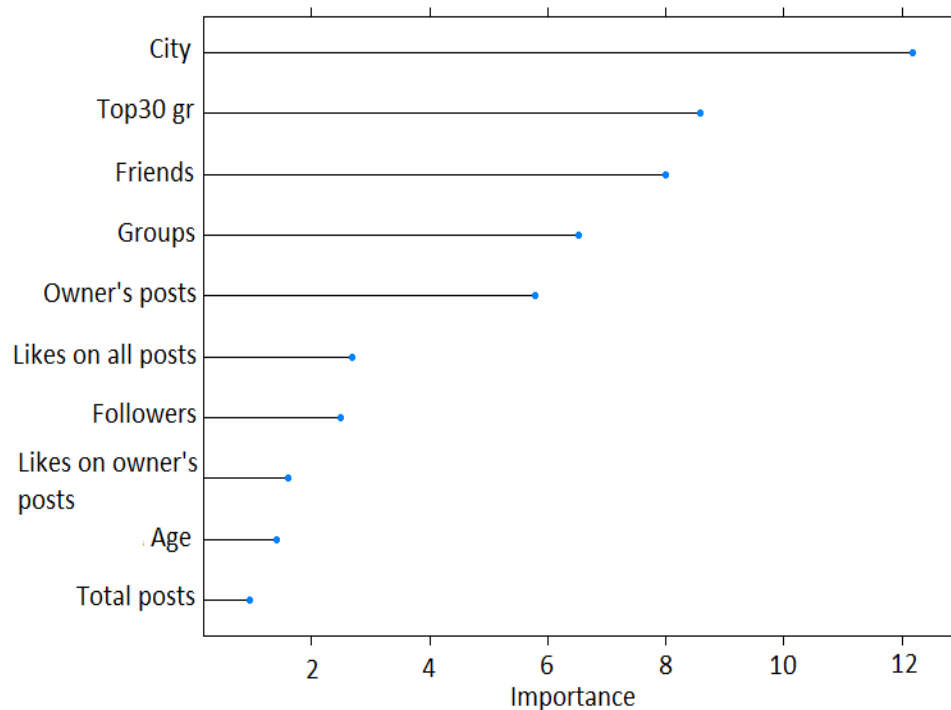
Results

Classification models' accuracies

Method	Team	Experts	PR	Finance	Law	Product	Rating	Grade
<i>Decision tree</i>	0.14	0.25	0.46	0.28	0.27	0.19	0.23	0.61
<i>Random forest</i>	0.1	0.21	0.45	0.3	0.33	0.25	0.2	0.52
<i>kNN</i>	0.12	0.22	0.45	0.2	0.35	0.15	0.2	0.49
<i>Gradient Boosting</i>	0.14	0.12	0.48	0.3	0.34	0.17	0.26	0.48

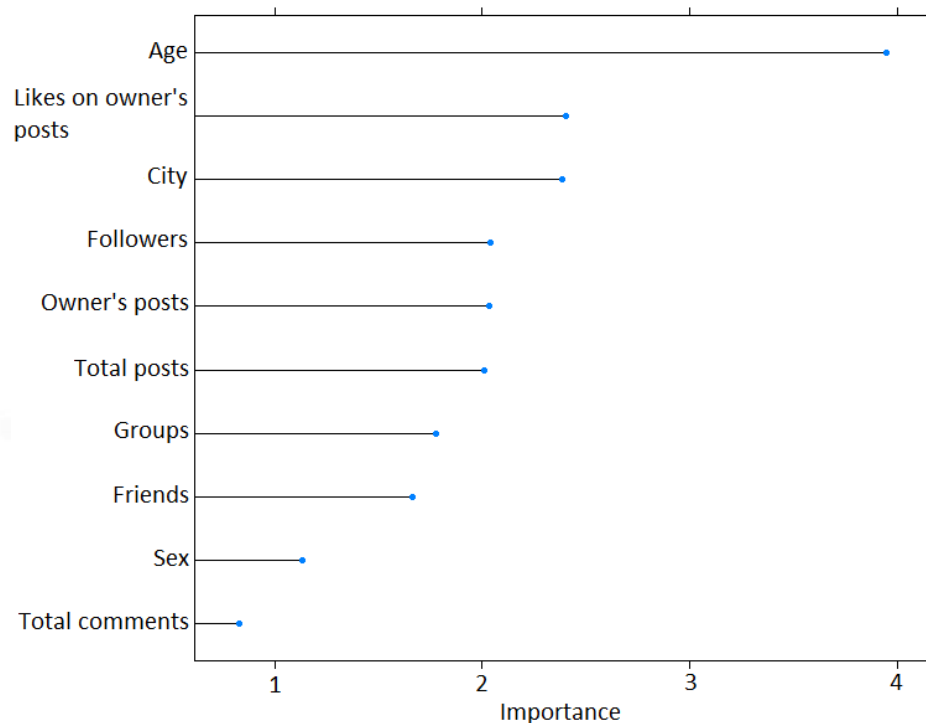
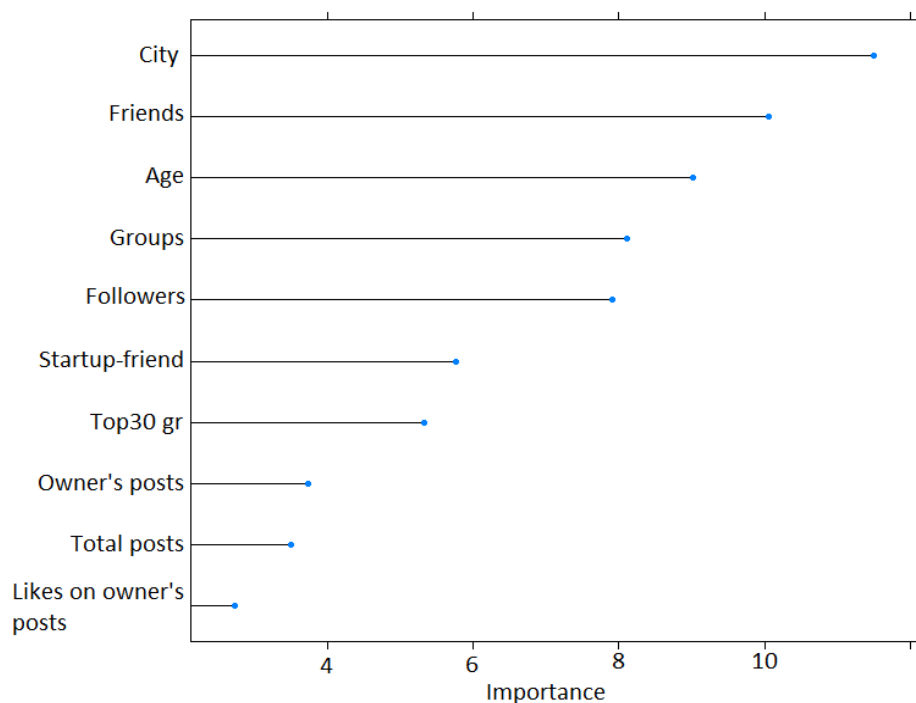
Results

Important features in predicting *Grade* (3 levels)



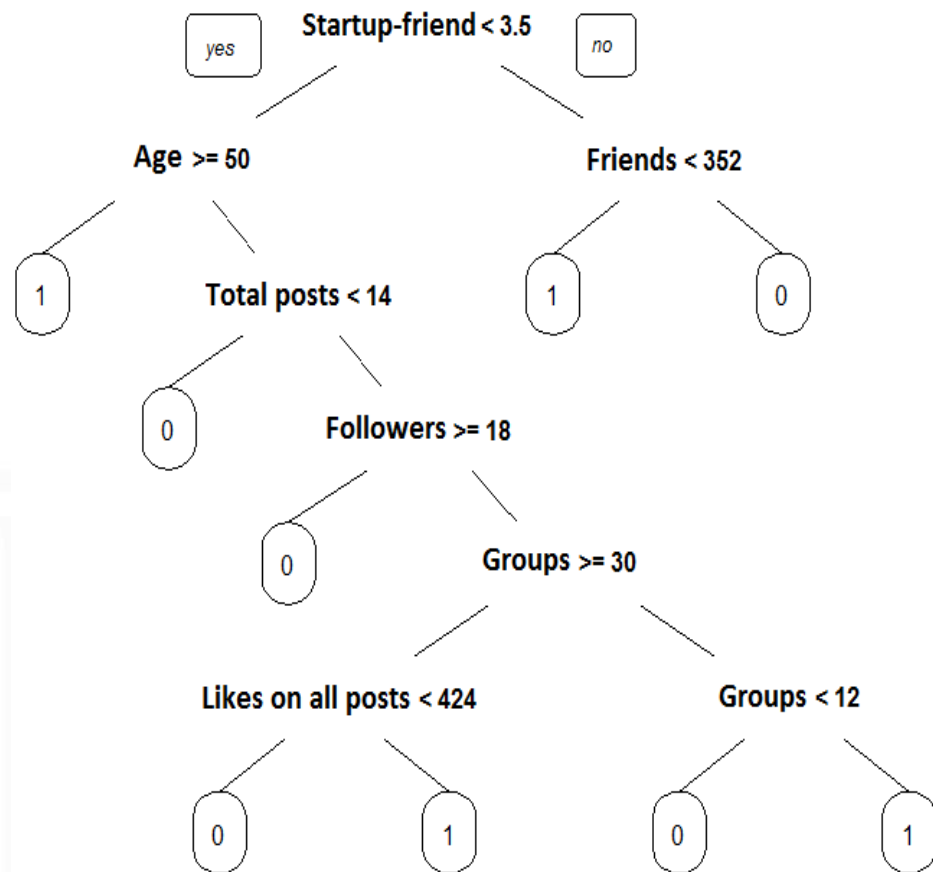
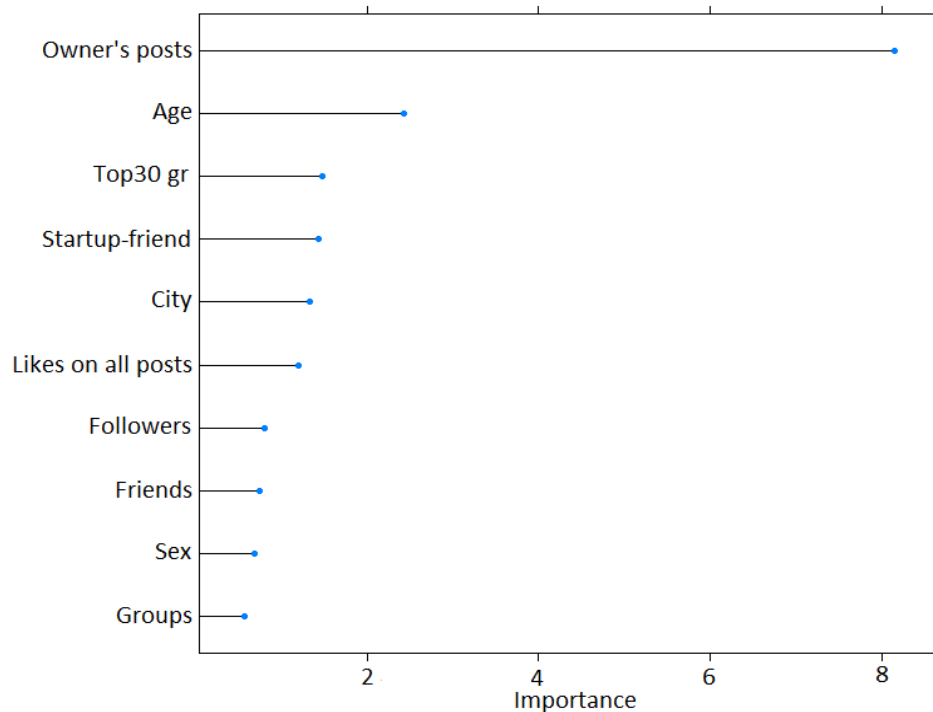
Results

Important features in predicting *Law* and *Team* ratings

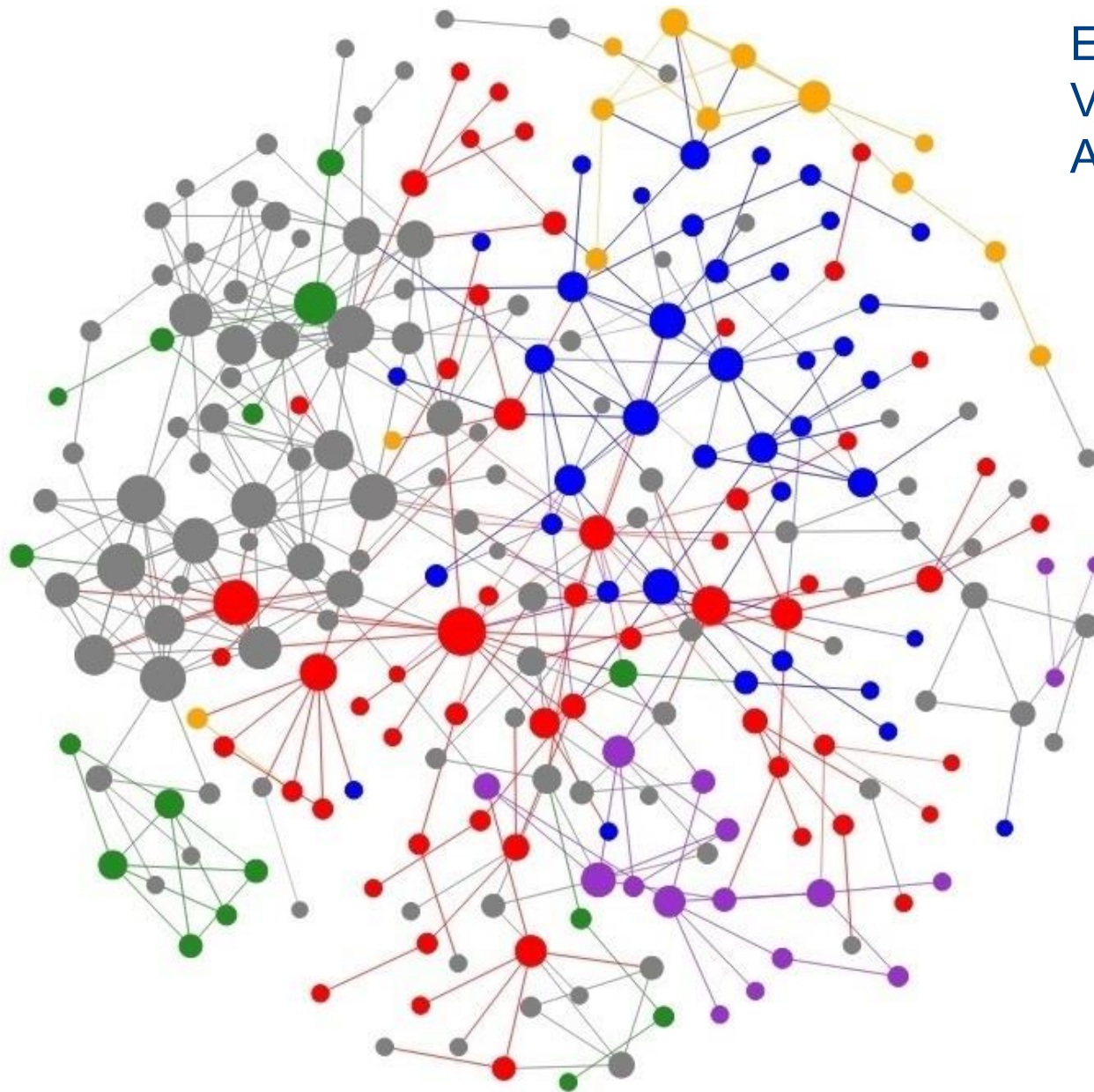


Results

Important features in predicting *PR* rating



Geographical distribution



Edge = 409

Vertex = 247

Assortativity (city) = 0.3613

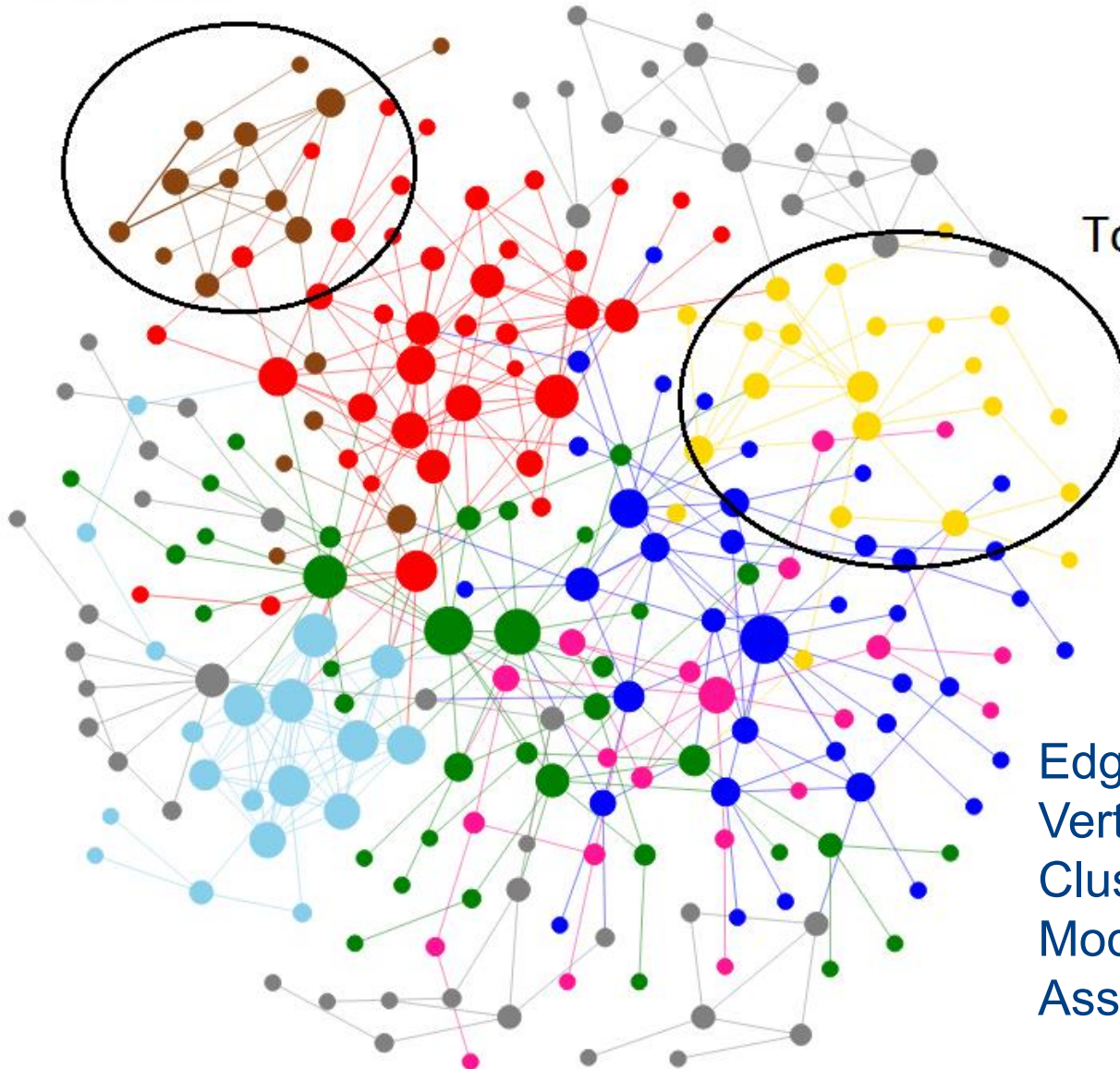
City:

- - Moscow
- - Saint Petersburg
- - Kazan
- - Tomsk
- - Astrahan
- - others

Clustering (fast greedy)

Astrahan

Tomsk



Edge = 405

Vertex = 241

Clusters = 8

Modularity = 0.74

Assortativity (degree) = 0.2



NATIONAL RESEARCH
UNIVERSITY

Thank you
for your attention!

Data problem

Imbalanced samples

Entrepreneurs	Random	% ratio
623	912584	0.068% / 99.932%
623	22580	2.68% / 97.32%

Way out: oversampling, undersampling,
SMOTE, ROSE

Descriptive statistics

Variable	Startupper, mean (SD)	Random user, mean (SD)	t Stat
<i>Friends</i>	310.34 (404.83)	51.09 (202.93)	15.96***
<i>Groups</i>	63.06 (100.24)	25.23 (90.645)	9.40***
<i>Followers</i>	112.72 (255.25)	27.77 (226.71)	8.29***
<i>All posts</i>	509.85 (1053.1)	119.60 (525.70)	7.74***
<i>Owner's posts</i>	352.61 (692.13)	95.14 (525.70)	7.75***
<i>All comments</i>	194.44 (776.36)	12.92 (132.65)	4.89***
<i>Owner's comments</i>	71.88 (284.15)	5.79 (59.46)	4.87***
<i>Likes on all posts</i>	929.71 (1950.0)	188.74 (881.25)	7.95***
<i>Likes on owner's posts</i>	892.73 (1911.2)	181.85 (861.26)	7.78***

Methodology

To test the **1st** hypothesis:

- *Student's t-test*
- *logistic regression*

To test the **2nd** hypothesis:

- *Spearman rank correlation coefficient*
- *classification algorithms*
- *a bit of network analysis*