Homophily and the Formation of Online Social Networks

Ruben Enikolopov, Maria Petrova, and Adam Szeidl

NES, UPF, and CEU

Social Diversity, Development and Stability, Moscow, October 2019

Overview

- Network structure of social interactions is important for real world outcomes
- Homophily (tendency to make connections with similar people) is a common property of interpersonal networks
  - leads to segregation
- Emergence of online social media can alter societal outcomes by providing people with new opportunities to connect

- How do decisions to join online social networks depend on offline interactions?
- Do online social networks promote or discourage segregation?

Overview

- Network structure of social interactions is important for real world outcomes
- Homophily (tendency to make connections with similar people) is a common property of interpersonal networks
  - leads to segregation
- Emergence of online social media can alter societal outcomes by providing people with new opportunities to connect
- How do decisions to join online social networks depend on offline interactions?
- Do online social networks promote or discourage segregation?

Overview

- Network structure of social interactions is important for real world outcomes
- Homophily (tendency to make connections with similar people) is a common property of interpersonal networks
  - leads to segregation
- Emergence of online social media can alter societal outcomes by providing people with new opportunities to connect

- How do decisions to join online social networks depend on offline interactions?
- Do online social networks promote or discourage segregation?

Overview

- Network structure of social interactions is important for real world outcomes
- Homophily (tendency to make connections with similar people) is a common property of interpersonal networks
  - leads to segregation
- Emergence of online social media can alter societal outcomes by providing people with new opportunities to connect

- How do decisions to join online social networks depend on offline interactions?
- Do online social networks promote or discourage segregation?

Overview

- Network structure of social interactions is important for real world outcomes
- Homophily (tendency to make connections with similar people) is a common property of interpersonal networks
  - leads to segregation
- Emergence of online social media can alter societal outcomes by providing people with new opportunities to connect

- How do decisions to join online social networks depend on offline interactions?
- Do online social networks promote or discourage segregation?

Homophily and Segregation

- Plentiful evidence of existence of homophily in offline (McPherson et al 2001; Jackson 2008) and online (Bakshy et al 2015) networks
- What affects segregation in networks?
  - Preference for same-type and biased matching both increase homophily in offline student networks (Currarini, Jackson and Pin 2009).
  - Random search for friends-of-friends can increase integration because friends-of-friends more representative (Bramoulle et al 2012).
- But:
  - Matching process is very different in online networks
  - Interactions between offline and online networks are not well understood
    - Bailey et al (2017, 2018) is a first step to bridging the gap

Homophily and Segregation

- Plentiful evidence of existence of homophily in offline (McPherson et al 2001; Jackson 2008) and online (Bakshy et al 2015) networks
- What affects segregation in networks?
  - Preference for same-type and biased matching both increase homophily in offline student networks (Currarini, Jackson and Pin 2009).
  - Random search for friends-of-friends can increase integration because friends-of-friends more representative (Bramoulle et al 2012).
- But:
  - Matching process is very different in online networks
  - Interactions between offline and online networks are not well understood
  - Bailey et al (2017) is a first step to understand this

Homophily and Segregation

- Plentiful evidence of existence of homophily in offline (McPherson et al 2001; Jackson 2008) and online (Bakshy et al 2015) networks
- What affects segregation in networks?
  - Preference for same-type and biased matching both increase homophily in offline student networks (Currarini, Jackson and Pin 2009).
  - Random search for friends-of-friends can increase integration because friends-of-friends more representative (Bramoulle et al 2012).
- But:
  - Matching process is very different in online networks
  - Interactions between offline and online networks are not well understood
    - Bailey et al. (JEP, 2018) is a first step in this direction

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

Online Networks and Segregation

- "Echo chambers":
  - People mostly get exposure to like-minded content (Sunstein 2002, 2017)
  - Limited exposure to cross-cutting content in Facebook mainly due to homophily in online friendship (Bakshy et al. 2015).
- Ambiguous effect on political polarization:
  - Online social networks promote polarization (Lelkes et al. 2017, Alcott et al. 2019, Mosquera et al. 2019)
  - Online social networks reduce polarization (Barbera 2017, Boxell et al. 2017)
- We look at online friendships in the presence of offline differences and focus on ethnic segregation

What we do

- Focus on **ethnic** segregation
- Explore friendship structure of Armenians in Russian online network Vkontakte and relate it to ethnic composition of cities.
- Develop a model of joint friendship choice online and offline
- Make predictions about structure of friendships online and offline
- Use data on users of Armenian heritage in VK, the leading Russian online social network, together with share of Armenians living in a city
  - Look at the online friendship choices by Armenians and non-Armenians
  - compare friends in the same and different cities
  - examine takeup of the network

What we do

- Focus on **ethnic** segregation
- Explore friendship structure of Armenians in Russian online network Vkontakte and relate it to ethnic composition of cities.
- Develop a model of joint friendship choice online and offline
- Make predictions about structure of friendships online and offline
- Use data on users of Armenian heritage in VK, the leading Russian online social network, together with share of Armenians living in a city
  - Look at the online friendship choices by Armenians and non-Armenians
  - compare friends in the same and different cities
  - examine takeup of the network

What we do

- Focus on **ethnic** segregation
- Explore friendship structure of Armenians in Russian online network Vkontakte and relate it to ethnic composition of cities.
- Develop a model of joint friendship choice online and offline
- Make predictions about structure of friendships online and offline
- Use data on users of Armenian heritage in VK, the leading Russian online social network, together with share of Armenians living in a city
  - Look at the online friendship choices by Armenians and non-Armenians
  - compare friends in the same and different cities
  - examine takeup of the network

What we do

- Focus on **ethnic** segregation
- Explore friendship structure of Armenians in Russian online network Vkontakte and relate it to ethnic composition of cities.
- Develop a model of joint friendship choice online and offline
- Make predictions about structure of friendships online and offline
- Use data on users of Armenian heritage in VK, the leading Russian online social network, together with share of Armenians living in a city
  - Look at the online friendship choices by Armenians and non-Armenians
  - compare friends in the same and different cities
  - examine takeup of the network

What we do

- Focus on **ethnic** segregation
- Explore friendship structure of Armenians in Russian online network Vkontakte and relate it to ethnic composition of cities.
- Develop a model of joint friendship choice online and offline
- Make predictions about structure of friendships online and offline
- Use data on users of Armenian heritage in VK, the leading Russian online social network, together with share of Armenians living in a city
  - Look at the online friendship choices by Armenians and non-Armenians
  - compare friends in the same and different cities
  - examine takeup of the network

Overview of the results

- Evidence that minorities are looking for online friendships with co-ethnics.
- Minorities are more likely to join online networks if they do not find sufficient number of co-ethnics offline.
- Implication: online social media can promote segregation

Overview of the results

- Evidence that minorities are looking for online friendships with co-ethnics.
- Minorities are more likely to join online networks if they do not find sufficient number of co-ethnics offline.
- Implication: online social media can promote segregation

Overview of the results

- Evidence that minorities are looking for online friendships with co-ethnics.
- Minorities are more likely to join online networks if they do not find sufficient number of co-ethnics offline.
- Implication: online social media can promote segregation

Model of online network formation

- Add pre-existing network and targeted search to CJP-style model.
- Agent of type i maximizes

$$U = \log(\gamma(m_s + n_s) + m + n) - \kappa t$$

where

  - In exogenous offline network m_s same type and m_d different type friends.
  - γ measures preference for same type.

- Find online friends n_s, n_d by choosing search time t and targeting ρ:

$$n_s = q_0 \cdot t + \rho q_0 \cdot t$$
$$n_d = (1 - q_0) \cdot t - \rho(q_0 + \theta/2\rho) \cdot t$$

where q_0 is share of same type among friends-of-friends.

  - Targeted search has effectiveness q_0 and convex cost.
  - No bias in matching process.

Solution

- **Proposition.** Denote $n_s/n = q_s$, then:

$$q_s = q_0 \frac{1 + (\gamma/\theta) q_0}{1 - q_0^2 \gamma^2/(2\theta)}$$

$$n = \frac{1 - q_0^2 \gamma^2/(2\theta)}{\kappa} - \frac{\gamma m_s + m}{1 + \gamma q_s}.$$

- Offline network m and opportunities q_0 jointly shape online network.
- **Predictions:**
  1. Substitution effect: $\partial n/\partial m_s < 0$ iff $\gamma > 0$.
  2. Targeting responds to opportunity: $q_s > q_0$ and $\partial q_s/\partial q_0 > 1$ iff $\gamma > 0$ and $\theta < \infty$.
  3. Substitution in entry: with entry cost, $\partial \text{entry}/\partial m_s < 0$, though with diminishing returns $\partial^2 \text{entry}/\partial m_s^2 > 0$
  4. All effects weaker if γ small.

Empirical Setting

- Information from VK – the main social network in Russia
  - Very similar to Facebook, but ten times more popular in Russia
- Get information on city, last name, and list of friends of the users of VK
  - All the data aggregated to the city level
  - Sample includes 625 cities with population above 20,000 excluding Moscow and Saint-Petersburg (as outliers).
- Classify as Armenians all the users who's last name ends in "-ian" or "-iants"
  - Classification validated using a dataset on victims of Great Terror in 1930's that contains name and ethnicity for more than a million people
- We assume that same city online friendship links mimic the same city offline friendship links

Summary Statistics

			N	Median	Mean	SD
Share of Armenians in the Ppopulation			625	0.31%	0.69%	1.42%
Share of Armenians among VK Users			625	0.43%	0.72%	1.12%
Share of Population with VK Accounts			625	59.01%	66.44%	32.25%
Share of Armenians with VK Accounts			623	84.60%	105.87%	100.40%
Armenian Users	Total # of Friends	Same city	622	28.49	31.80	15.81
		Different cities	624	35.34	40.34	27.53
	Share of Armenian Friends (weighted by total number of friends)	Same city	622	4.54%	5.79%	4.92%
		Different cities	624	10.10%	10.70%	5.53%
Non-Armenian Users	Total # of Friends	Same city	625	28.38	31.58	15.27
		Different cities	625	37.16	39.17	10.76
	Share of Armenian Friends (weighted by total number of friends)	Same city	625	0.41%	0.66%	0.99%
		Different cities	625	0.58%	0.69%	0.45%

- Inbreeding homophily for Armenians, especially among different-city friends.

Number of Friends

	Armenian Users Total # of Friends		Non-Armenian Users Total # of Friends	
	Same city	Different cities	Same city	Different cities
	(1)	(2)	(3)	(4)
Share of Armenians in population	31.353	-169.686**	25.032	26.787
	[26.386]	[67.063]	[24.855]	[18.885]
Controls	Yes	Yes	Yes	Yes
Observations	622	624	625	625

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in brackets are adjusted by clusters within regions. Unit of observation is a city. Controls include population (5th polynomial), number of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older years, in each city according to 2010 Russian Censu; share of population with higher education overall according to 2002 Russian Census and separately in each of the age cohorts according to 2010 Russian Census, dummy for regional and county centers, distances to Moscow and St Petersburg, log (average wage), share of people with higher education in 2002, internet penetration in 2011, log (Odnoklassniki users in 2014), federal region fixed effects.

- Negative different-city coefficient suggests substitution between online and offline Armenian friends and $\gamma > 0$. Consistent with the model's first prediction.
- Generally smaller effects for non-Armenians may be explained by declining γ once some threshold level of co-ethnic friendship is reached, e.g. when "safety network" is maintained.

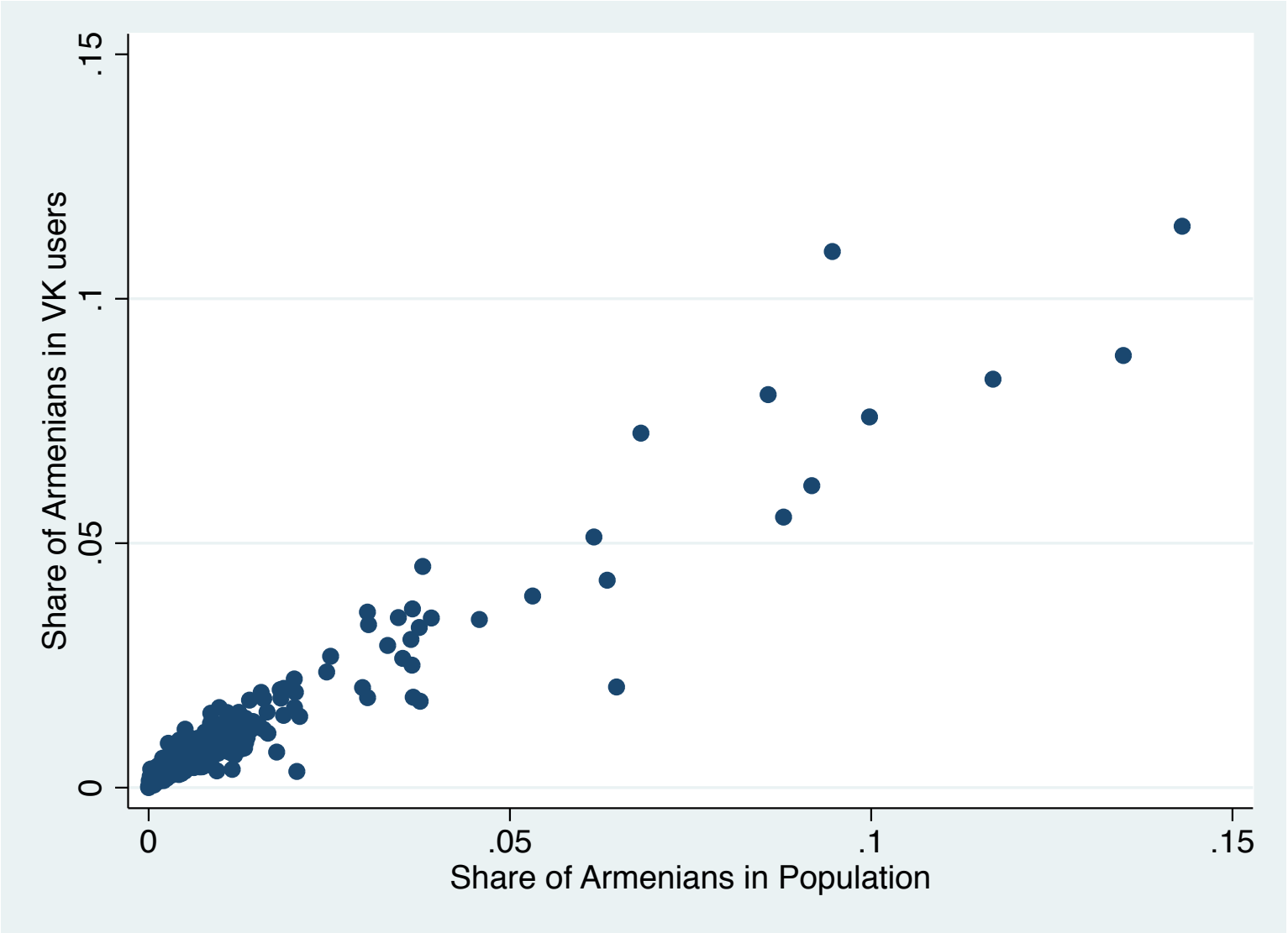
Share of Armenian Friends

	Armenian Users		Non-Armenian Users	
	Share of Armenian Friends (weighted by total number of friends)		Share of Armenian Friends (weighted by total number of friends)	
	Same city	Different cities	Same city	Different cities
	(1)	(2)	(3)	(4)
Share of Armenians in population	2.601***	1.530***	0.640***	0.191***
	[0.148]	[0.226]	[0.023]	[0.015]
Controls	Yes	Yes	Yes	Yes
Observations	622	624	625	625

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in brackets are adjusted by clusters within regions. Unit of observation is a city. Controls include population (5th polynomial), number of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older years, in each city according to 2010 Russian Censu; share of population with higher education overall according to 2002 Russian Census and separately in each of the age cohorts according to 2010 Russian Census, dummy for regional and county centers, distances to Moscow and St Petersburg, log (average wage), share of people with higher education in 2002, internet penetration in 2011, log (Odnoklassniki users in 2014), federal region fixed effects.

- Larger different-city coefficient for Armenians is consistent with targeted search.

Takeup



Most points seem to lie below the diagonal

Take-up

	Share of Armenians with VK Accounts		Share of non-Armenians with VK Accounts	
	(1)		(2)	
Share of Armenians in population	-38.897**	-59.030***	2.952	-5.318
	[15.022]	[15.116]	[2.094]	[3.558]
Share of Armenians in population squared	323.019**	553.223***	-5.375	66.415**
	[130.834]	[169.510]	[20.874]	[30.807]
Log (SPbSU students), same cohort as VK founder		0.067		-0.015
		[0.088]		[0.025]
Log (SPbSU students), same cohort as VK founder x		17.283*		10.896**
Share of Armenians in population		[9.866]		[4.514]
Log (SPbSU students), same cohort as VK founder x		-185.940*		-83.107**
Share of Armenians in population squared		[94.153]		[33.041]
Controls	Yes		Yes	
Observations	623		625	

Consistent with substitution effect, takeup is negatively related to offline share of Armenians, with diminishing returns.

The effect is smaller in places with a larger initial shock to aggregate network penetration, consistent with a model.

Share of Friends from Other Cities

	Armenian Users Share of Friends from Different Cities		Non-Armenian Users Share of Friends from Different Cities	
	Armenian Friends	Non-Armenian Friends	Armenian Friends	Non-Armenian Friends
	(1)	(2)	(3)	(4)
Share of Armenians in population	-2.571***	-1.452***	-3.511***	-0.343
	[0.421]	[0.388]	[0.649]	[0.221]
Controls	Yes	Yes	Yes	Yes
Observations	622	624	625	625

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in brackets are adjusted by clusters within regions. Unit of observation is a city. Controls include population (5th polynomial), number of people aged 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older years, in each city according to 2010 Russian Censu; share of population with higher education overall according to 2002 Russian Census and separately in each of the age cohorts according to 2010 Russian Census, dummy for regional and county centers, distances to Moscow and St Petersburg, log (average wage), share of people with higher education in 2002, internet penetration in 2011, log (Odnoklassniki users in 2014), federal region fixed effects.

Presumably more offline Armenian friends in the same city implies lower share of Armenian friends from other cities.

For non-Armenian friends, consistent with substitution effect.

Conclusions

- Model suggests two forces shaping search for online friends:
  - substitution for offline friendships
  - opportunity of targeted search
- Evidence on Armenians in Vkontakte shows:
  - Inbreeding homophily stronger in online than offline;
  - Substitution between same-type online and offline friends;
  - Biased matching perhaps driven by targeted search;
  - More entry when fewer Armenians offline.
- Our current interpretation: incentive to find co-ethnics is important determinant of online networks and increases segregation.
- More generally, social media can promote segregation of interpersonal interactions and prevent assimilation.