

Network Analysis of Media Users

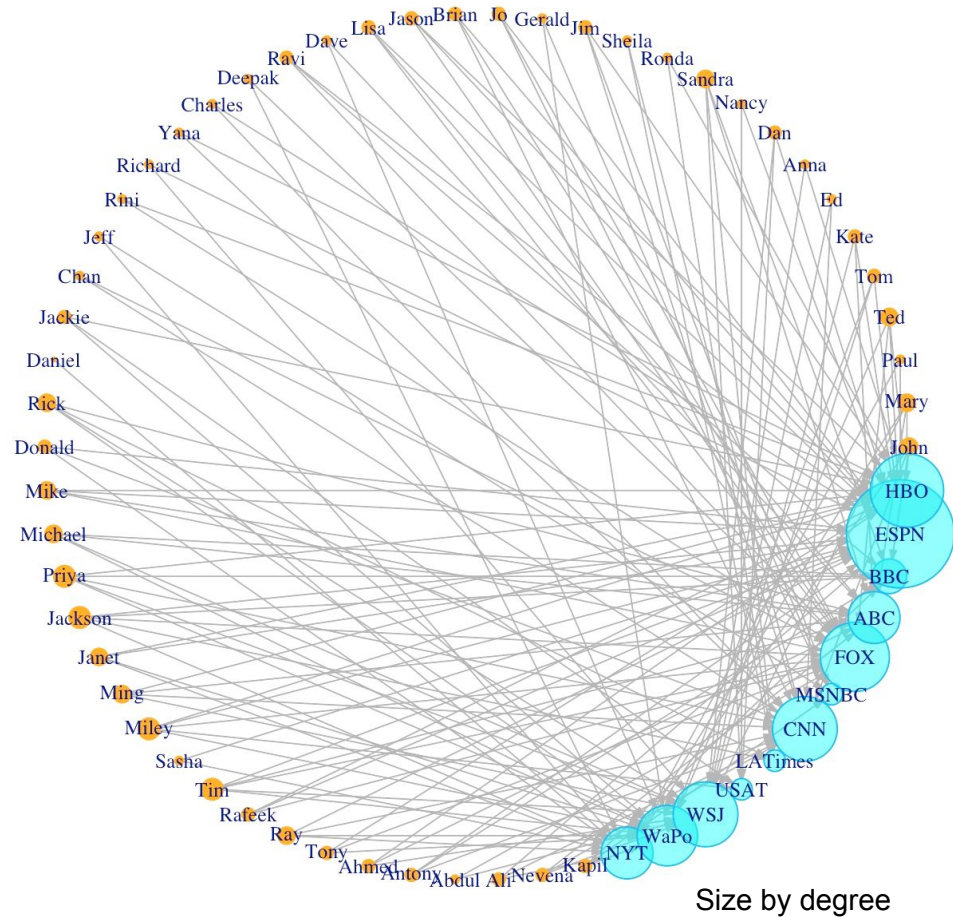
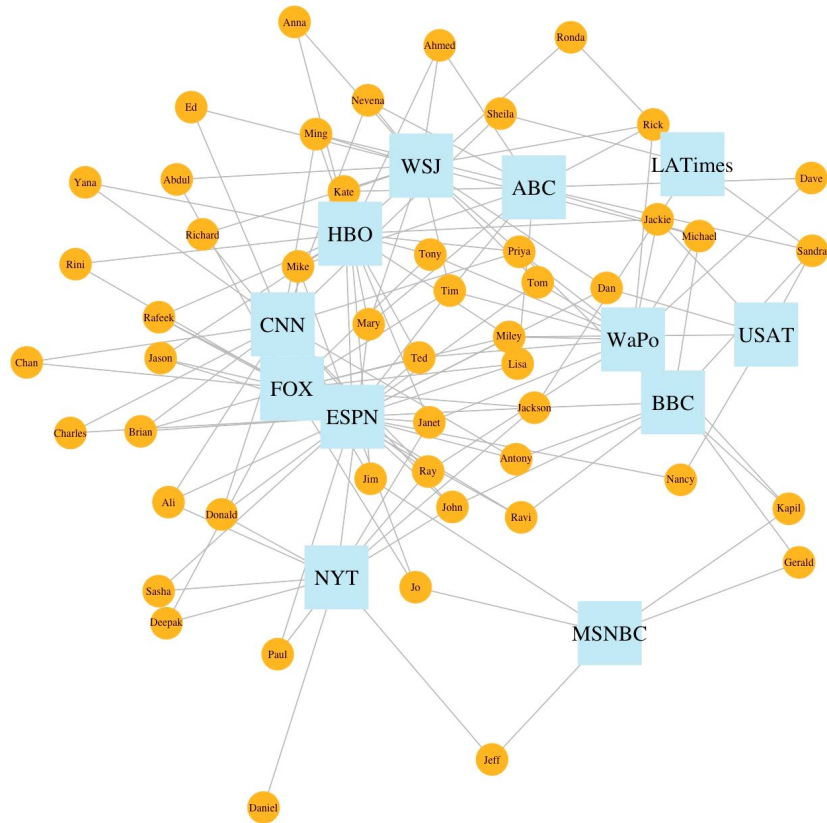
Group 6

Ruoyun Deng
Naveen Ithikkat
Rongchu Zhao

Contents

- Background
 - About Data
 - Our Questions
- Two-Mode Network
 - Full network
 - Subgroup network according to different attributes
 - ERGM with external network
- Conversion
- Converted One-Mode Network
 - ERGM
- Conclusion

Views of the network



About the Data

- 'Two-mode' network - i.e, the nodes in the network may represent users (persons) or media channels (like CNN, FOX, etc.)
- Original dataset from [POLNET 2016 Workshop](#)
 - Nodes - 30 (10 media, 20 users), Edges - 31
- Conducted a survey within friends about their choice of media
 - Final dataset: Nodes - 62 (12 media, 50 users), Edges - 149
- Attributes
 - Media: Media Name and Type (TV or News websites)
 - Users: Gender, Age and Marital Status

Our questions

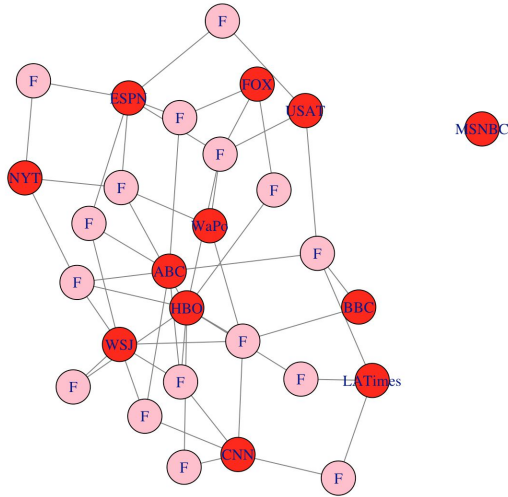
- Which are the most popular channels in the network overall?
- How many channels does a user consume on average?
- Based on gender, age, or marital status, does the choice of channels vary among different groups of users?
- Based on gender, age, or marital status, is there a specific group that consumes more or less channels on average?

What can we say about the ‘full’ network?

- The 3 most popular channels in the network are:
 - ESPN
 - HBO
 - FOX
- A user in the network consumes about 3 channels on average - 2 of them being TV channels and 1 news website.

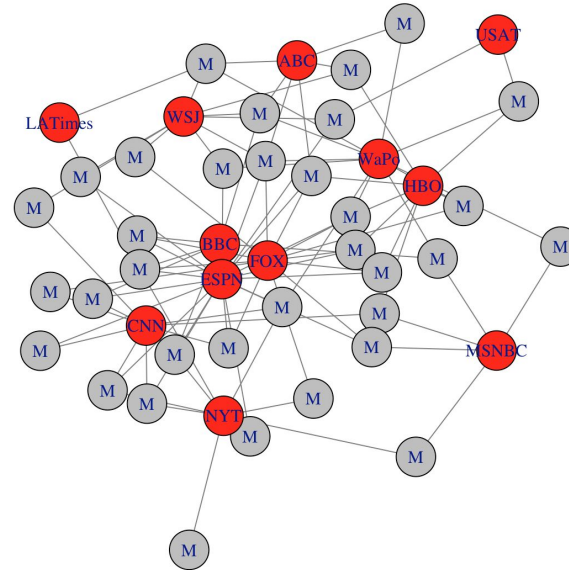
User behavior by gender

Female Network



Top channels: HBO, WSJ, ABC

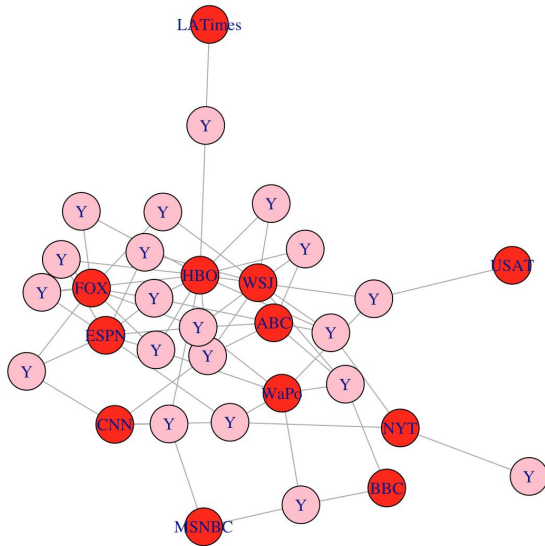
Male Network



Top channels: ESPN, FOX, WaPo

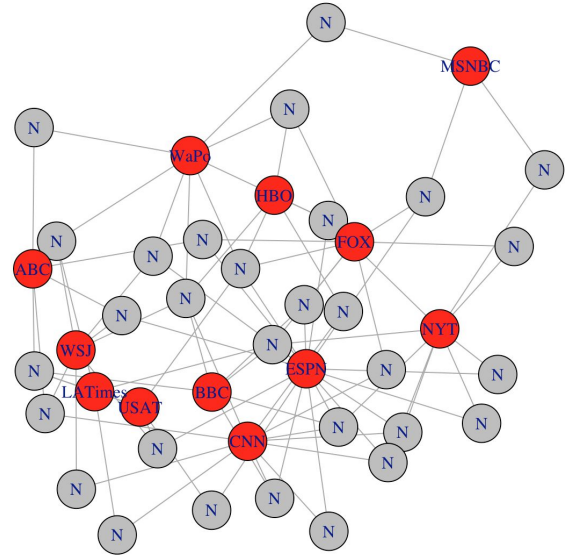
User behavior by marital status

Married Users



Top channels: HBO, WSJ, FOX

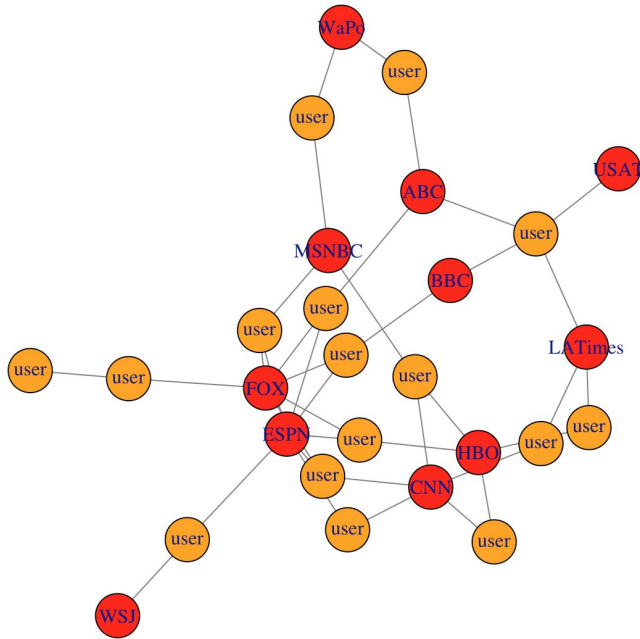
Unmarried Users



Top channels: ESPN, CNN, NYT

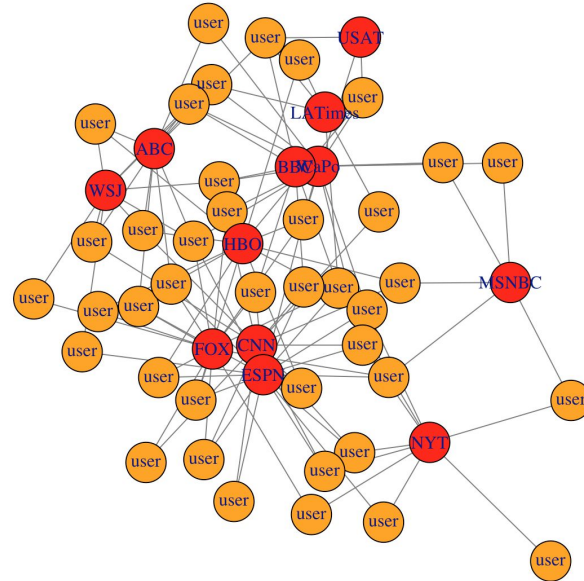
User behavior by age

31-40 age group



Top channels: ESPN, FOX, CNN

31-70 age group



Top channels: ESPN, FOX, HBO

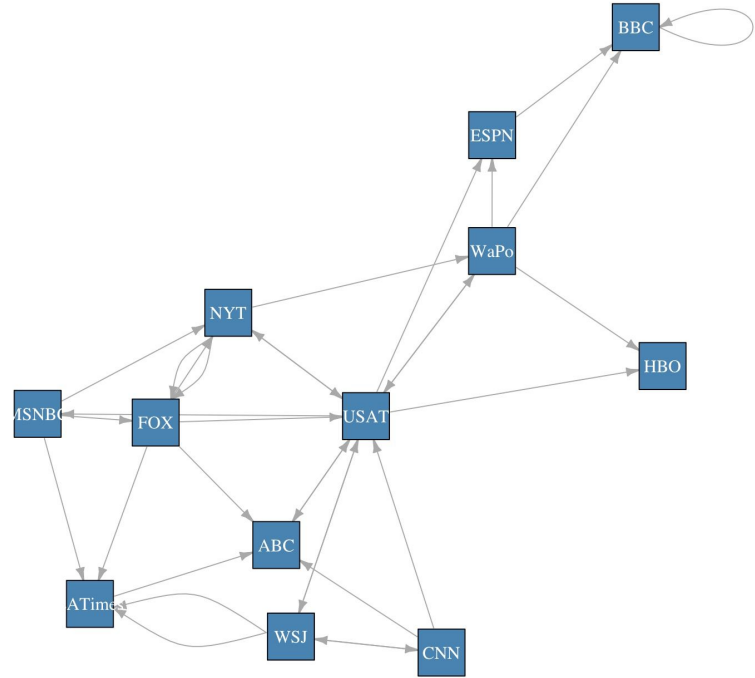
Summary of findings

- The most popular channels among males and females are very different.
- The most popular channels among married and unmarried users are very different.
- Popular channels by age group are NOT very different.
- The average number of channels consumed by users is about 3 regardless of the user's age, gender or marital status.

ERGM (with additional data)

Additional data about interactions (eg. hyperlinks / citations) between media channels is also available via kateto.net.

Using this additional data along with the original dataset, we are able to run an ERGM model for the new network.



ERGM on 2-mode network

The log-odds of two actors having a tie is:

$$-2.8 \times \{\text{change in the number of ties}\} + 0.67 \times \{\text{change in number of triangles}\}$$

- For a tie that will create no triangles, the log-odds changes by : (-2.8). (p=0.06)
- For a tie that will create one triangle, the log-odds changes by : (-2.8 + 0.67 = -2.13) (p=0.11)
- For a tie that will create two triangles, the log-odds changes by : (-2.8+0.67× 2= -1.46) (p=0.19)

**Note* Apart from finding probabilities of actors having a tie in the network, ERGM may not help with further interpretations(Homophily) in the case of 2-mode network.*

=====
Summary of model fit
=====

Formula: `mu.links.net ~ edges + triangle`

Iterations: 4 out of 20

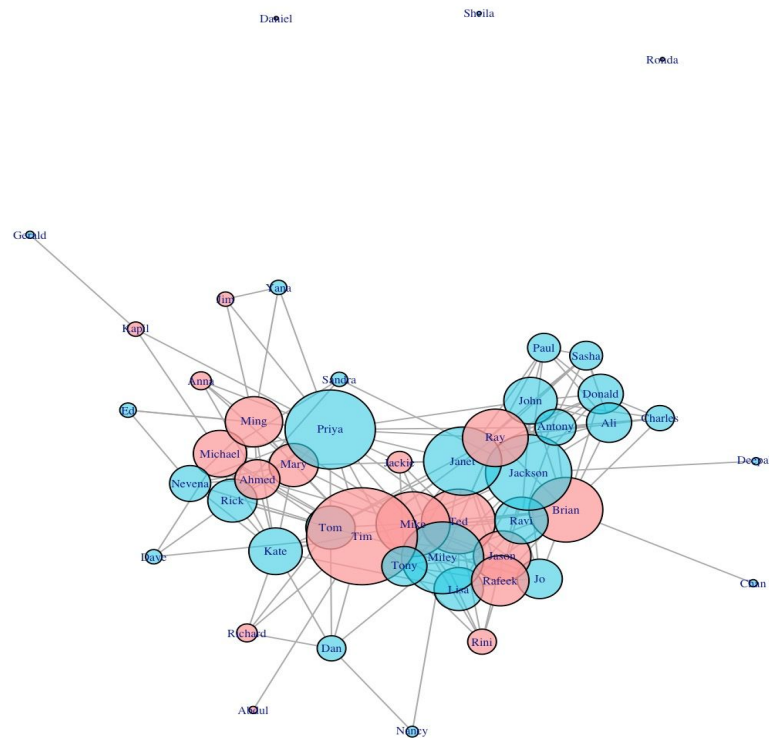
Monte Carlo MLE Results:

	Estimate	Std. Error	MCMC %	p-value
edges	-2.7695	0.1246	1	<1e-04 ***
triangle	0.6731	0.1176	1	<1e-04 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance:	2621	on 1891	degrees of freedom
Residual Deviance:	1152	on 1889	degrees of freedom

AIC: 1156 BIC: 1167 (Smaller is better.)



ERGM on 1-mode network

In the converted one-mode network, attributes of users did not have a significant result

Only p-value of marital status1. <0.05

Marital status is binary variable (0 for unmarried and 1 for married), so log-odds of two users who is married sharing media will change by (0.44) and the possibility is 61%

=====

Summary of model fit

=====

Formula: user.net ~ edges + nodematch("user.age") + nodematch("user.gender") + nodematch("marital.status", diff = T)

Iterations: 4 out of 20

Monte Carlo MLE Results:

	Estimate	Std. Error	MCMC %	p-value
edges	0.4232	0.1080	0	<1e-04 ***
nodematch.user.age	-0.5237	0.4067	0	0.1981
nodematch.user.gender	0.1199	0.1204	0	0.3195
nodematch.marital.status.0	0.1117	0.1310	0	0.3941
nodematch.marital.status.1	0.4407	0.1817	0	0.0154 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 1698 on 1225 degrees of freedom
Residual Deviance: 1590 on 1220 degrees of freedom

AIC: 1600 BIC: 1625 (Smaller is better.)

Conclusion

- Using network analysis, we can gain useful insights into media consumption patterns for users in a network.
- This information can be used to identify specific groups of customers to understand their behavior, target ads to specific groups, etc.
- A richer understanding of the network could be obtained if data about links between users is also included.

Thank You!