Modeling Relationship Strength using Phone Data

Emanuelle Alemar, Jonathan Che, Joshua Core, Mishal McNeill
Carnegie Mellon University, Pittsburgh, Pennsylvania

Introduction

Police often recover suspects' phones during criminal investigations. Manually examining phone data to uncover relationships between individuals, however, is both difficult and time-consuming.

In our project, we modeled and predicted relationships between individuals based only on their call and text logs. In particular, we focused on friend and couple relationships. Successful extensions of our work could be of great help to police in their investigations.

Research Questions

- What call and text information is useful for determining whether two individuals are friends?
- Can we predict whether two individuals are a couple?

Materials and Methods

Data Collection

- We used cellphone data collected by MIT's Human Dynamics Lab
- Data were collected from 200 participants in residential living community from March 2010 – September 2011
- Data contains information about call/text activity:
  - Number of calls/texts in/out between the individuals
  - Proportion of calls/texts in/out between the individuals
  - Number of people the first individual is in contact with
  - Total calls/texts made by the first individual

Friend Pair Dataset:

- Data were labeled by assigned friend score (0 is lowest, 7 is highest) or strength of friendship (0-4 is weak, 5-7 is strong)
- We used logistic regression and principal component analysis (PCA) to infer significant predictors of friendship strength

Couple Pair Dataset:

- Data were labeled by whether the pair of individuals are a couple
  - Only 36 couple relationships out of 6,000+ relationships
- We used logistic regression and PCA to predict couple status

Text and Call Relationships

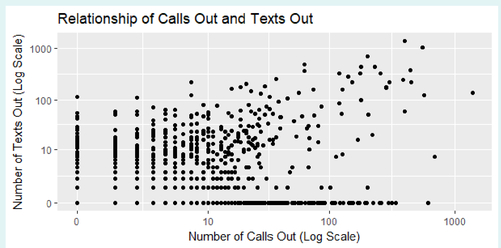


Figure 1: Scatterplot of calls out vs. texts out in a given relationship

In general, calling someone more is associated with texting them more as well. Notably, though, a significant number of relationships do not involve any texts at all.

Friend Score Relationships

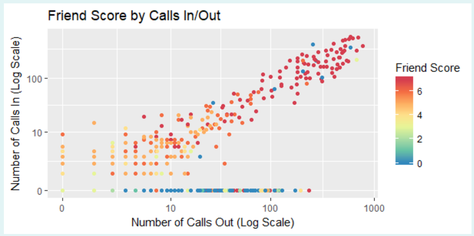


Figure 2: Relationships with more calls in/out have higher friend scores

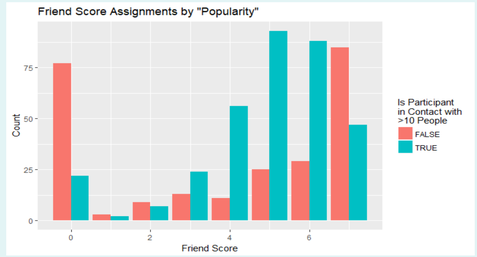


Figure 3: Participants in contact with many people tend to assign friend scores in the 4-6 range, while participants in contact with fewer people tend to assign friend scores of either 1 or 7.

Friend Score Modeling

We train logistic regression models to gain further insight into some of the more complex call/text behaviors associated with friendship.

Model 1: Strong Friendship ~ Is Popular + Proportion of Texts Out + (Is Popular * Proportion of Texts Out)

Variable	Estimate	P-Value
Is Popular	0.137	0.507
Proportion of Texts Out	4.566	0.007
Is Popular * Proportion of Texts Out	72.054	0.003

For participants in contact with many others, having a high proportion of texts out to a specific person is a much stronger indicator of a strong friendship than it is for participants in contact with fewer others.

Model 2 (PCA): Participant ID + PC1 + PC2 + PC3

Variable	Estimate	P-Value
PC1	0.470	0.00004
PC2	-4.20	0.001
PC3	1.66	0.0001

The three principal components can be interpreted as follows:

1) **Volume of texts and calls between the pair of people**
2) **Each individual's propensity to text rather than call**
3) **Volume of texts, rather than calls, between the pair of people**

In general, calling and texting more is associated with an increased odds of being strong friends. While individuals who text a lot tend to have weaker relationships in general, texting someone a lot is indicative of a strong friendship.

Couple Pair Relationships

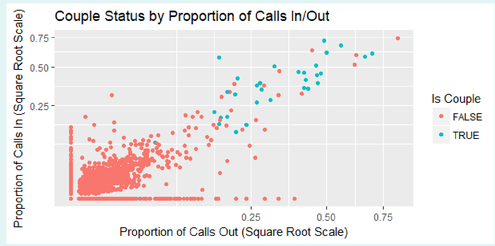


Figure 4: Couples tend to have a higher proportion of calls in and out with each other

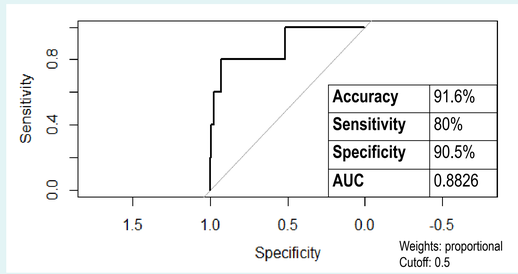
Proportion of total calls in and calls out can effectively separate couples from non-couples. In fact, a tree model built only on those two variables shows decent performance (~50% Sensitivity, ~99% Specificity).

Couple Pair Modeling

Model 3: Is Couple ~ (Variables in table)

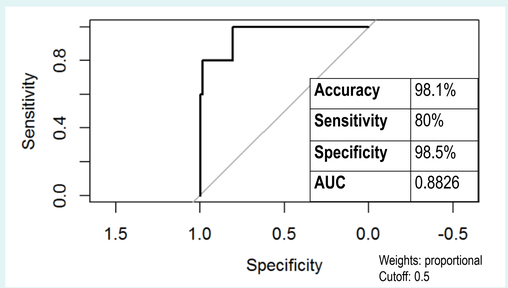
Variable	Estimate
Intercept	-3.52
Calls In-Night	-2.14
Calls In-Morning	-1.37
Calls In-Weekend	-2.49
Calls Out-Morning	3.96

SMS In-Night	2.15
SMS In-Morning	-7.03
SMS Out-Morning	6.28
Prop. Of Calls In	221.47
Prop. Of Calls Out	296.15
Prop. Of SMS Out	179.71



Model 4 (PCA): Is Couple ~ PC1 + PC2 + PC3

Variable	Estimate
Intercept	-1.89
PC1	0.288
PC2	0.49
PC3	0.216



Discussion

In our project, we used cell-phone call and text logs to model relationships between different people. A number of general trends emerged; for example, friend relationships that involved more calls and texts tended to be reported as stronger relationships.

We also found that the individuals who are in contact with many people seem to have more medium-strength relationships. The strength of these "popular" individuals' relationships is also more heavily related to their text activity.

Finally, we trained models to predict whether a given relationship is a couple relationship. Weighted logistic regression models performed decently well on the data, achieving high sensitivity while not predicting too many false positives. Interestingly, PCA helps separate couples from non-couples, which may indicate that couples just have significantly more contact with each other in some sense.

Limitations/Next Steps

Limitations:

- Dependencies due to network structure of data
- Unexplained errors/discrepancies in data collection process

Next Steps:

- Network modeling to account for dependencies
- Community detection

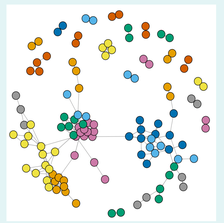


Figure 5a: Network graph of strong friend relationships, colored by detected communities.

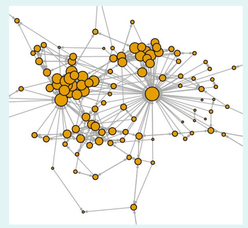


Figure 5b: Network graph of friend pair data, node size proportional to degree. Some individuals are contacted much more than others.

Conclusion

- Overall volume of calls and texts in a relationship can account for a significant portion of the strength of the relationship.
  - Individuals in contact with many others tend to contact more medium-level friends.
  - People disproportionately text their closest friends more.
- Couple relationships are easier to identify than friendships.
  - A high proportion of calls in/out is a good indicator of a couple relationship.

If there are any further questions, feel free to contact Jonathan Che at jche18@amherst.edu

Acknowledgements

We would like to thank our TA, Ciaran Evans, for all of his assistance, as well as the CMU Statistics Department, CSAFE, and NIST.