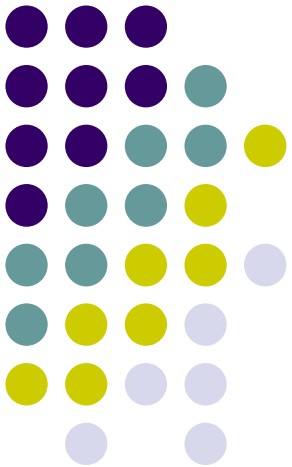
Advanced Computer Graphics
CS 525M: Social Sensing for Epidemiological Behavior Change

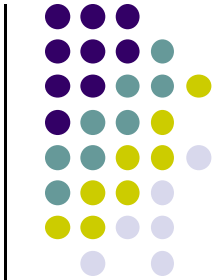
Zahid Mian

Computer Science Dept.

Worcester Polytechnic Institute (WPI)

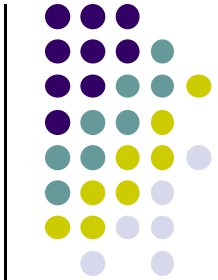


Epidemiology



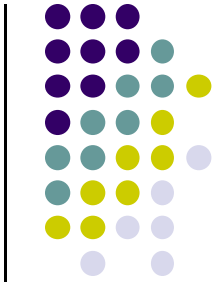
- Study of the …
  - Patterns
  - Causes
  - Effects of health and disease conditions
- … in defined populations
- Study of the causes and transmission of disease within a population

Paper Outline



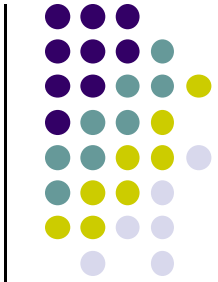
- The Problem
- Experiment/Infrastructure
- Results
  - Basic Patterns
  - K-Clustering Classification
  - PSI Algorithm
- Conclusion

What's the Problem?



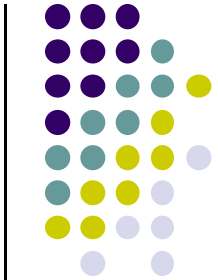
- "Understand how individual behavior patterns are affected by physical and mental health symptoms."
- Can cell phones be used to detect an outbreak of diseases? (ubiquitous computing)
  - Based on co-location of devices

Some Behavioral Scenarios



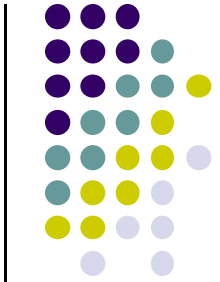
- Introverts, isolates, and persons lacking social skills may also be at increased risk for both illness behaviors and pathology.
- Stress depletes local immune protection, increasing susceptibility to colds and flu.
- Psychological disturbances could develop in response to frequent illness.

The Experiment



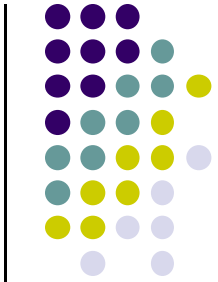
- Residents of an undergrad dorm (Feb to Apr 2009).
  - Data secured, anonymized
- Those immunized for influenza, filtered out (baseline survey)
- Individuals surveyed daily for symptoms of contagious diseases
  - common colds, influenza & gastroenteritis.
  - Phone disabled if survey not completed
- Characteristic changes in behavior when sick
  - total communication
  - communication patterns with respect to time of day
  - diversity of their network.

Infrastructure



- Device Selection
  - Client based on Windows Mobile 6.x devices
  - Supported devices featured WLAN,EDGE and SD Card storage
- Data
  - Logged Call and SMS details every 20 minutes
  - missed calls and calls not completed
- Server
  - Post-processing of logs

Mobile Sensing Platform



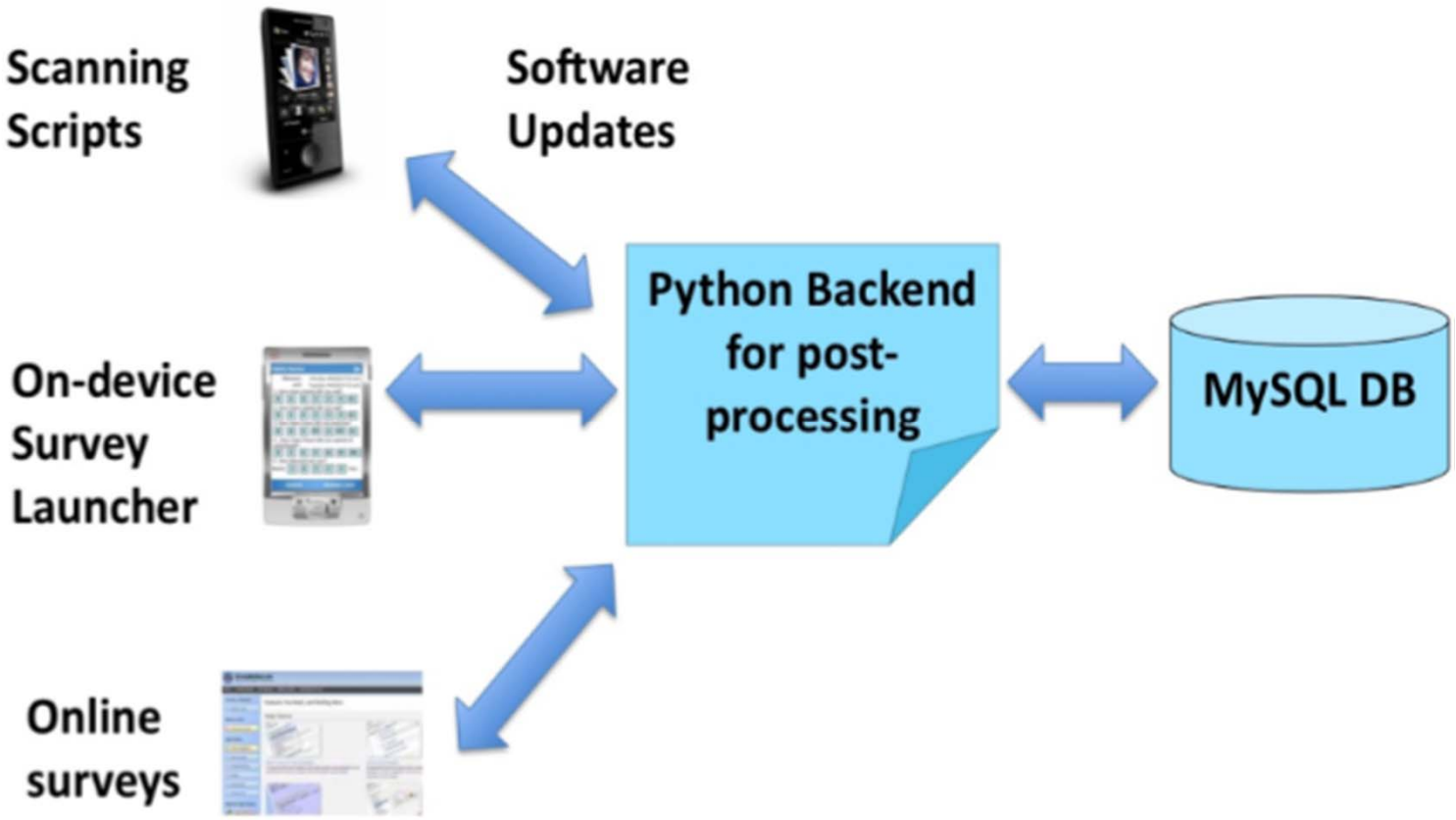
- Proximity Detection (Bluetooth):
  - looked for other Bluetooth-enabled devices
- Approximate Location (WLAN AP)
  - Determine Location based on the AP (over 55 APs available in building)
- Battery Impact
  - Windows Phones notoriously bad for battery
  - Limit periodic scanning

Daily Survey

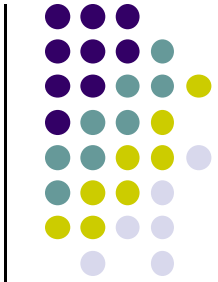
Table 1. Symptom Survey Questionnaire. All questions were Yes/No responses

Survey Question (as shown on mobile phone)
Do you have a sore throat or cough?
Do you have a runny nose, congestion or sneezing?
Do you have a fever?
Have you had any vomiting, nausea or diarrhea?
Have you been feeling sad, lonely or depressed lately?
Have you been feeling stressed out lately?

Platform Architecture

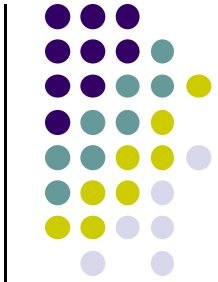


Terms



- Total Communications: phone calls + SMS
- Late Night/Early Morning: 10 pm – 9 am
- Communication Diversity: unique individuals within 48-hour period
- Entropy: Amount of disorder or randomness
- Physical Proximity *Entropy* with Other Participants: # of times remote device scanned divided by total scanned devices in 48 hours

Terms



- Physical Proximity Entropy with Other Participants Late Night and Early Morning
- Physical Proximity Entropy for Bluetooth Devices Excluding Experimental Participation
  - “familiar strangers” Bus Stop, classroom, etc.
- WLAN Entropy based on University WLAN APs
  - Only University APs are considered
- WLAN Entropy based on external WLAN APs

Behavioral Effects of Low Intensity Symptoms (Runny nose)

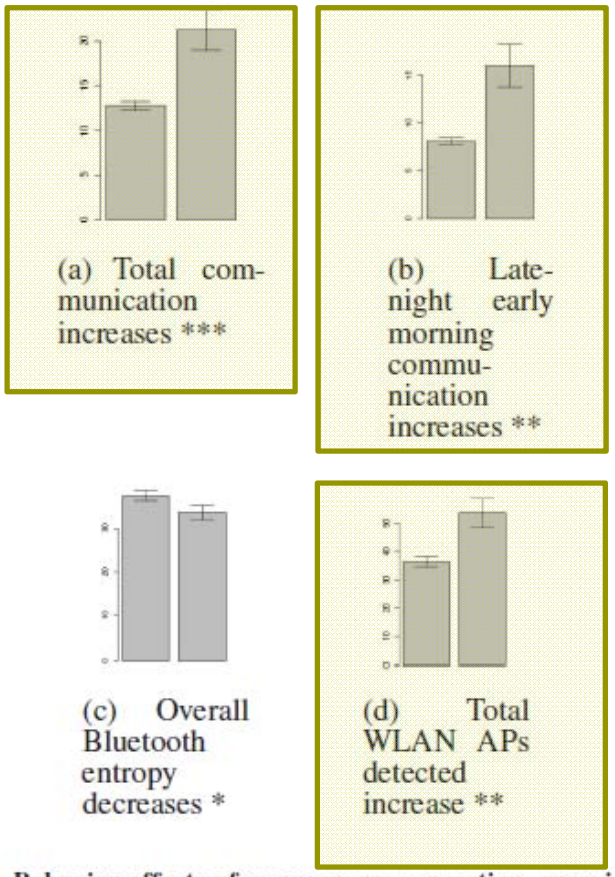
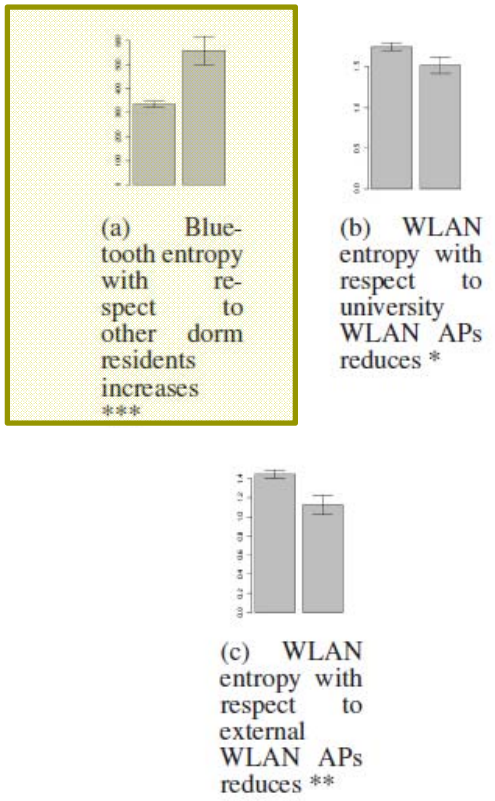


Figure 1. Behavior effects of runny nose, congestion, sneezing symptom, n=587/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

- Increased total communication
- Increased late-night early morning communications
- Increased WLAN APs

Behavior Effects of Low-Intensity Symptoms (Sore Throat)



(a) Bluetooth entropy with respect to other dorm residents increases ***

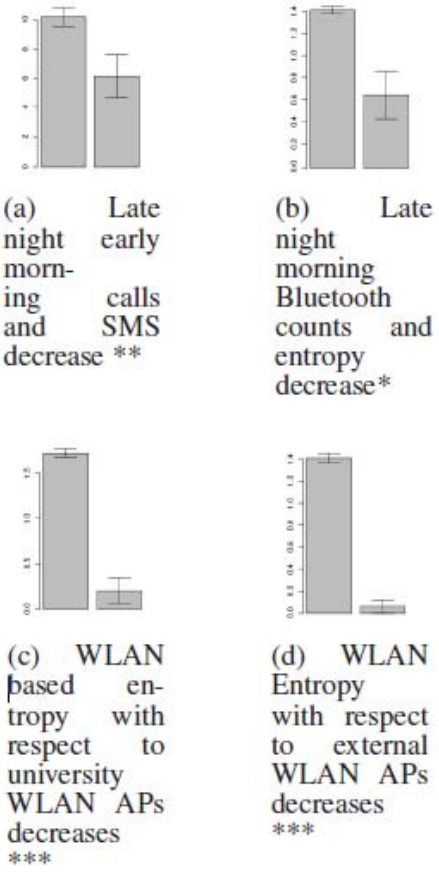
(b) WLAN entropy with respect to university WLAN APs reduces *

(c) WLAN entropy with respect to external WLAN APs reduces **

Figure 2. Behavior effects of sore throat and cough symptom, n=393/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

- Bluetooth entropy with respect to others increases
  - counter-intuitive
  - Spending time indoors?

Behavior Effects of Higher-Intensity Symptoms (Fever and Influenza)



(a) Late night early morning calls and SMS decrease **

(b) Late night morning Bluetooth counts and entropy decrease*

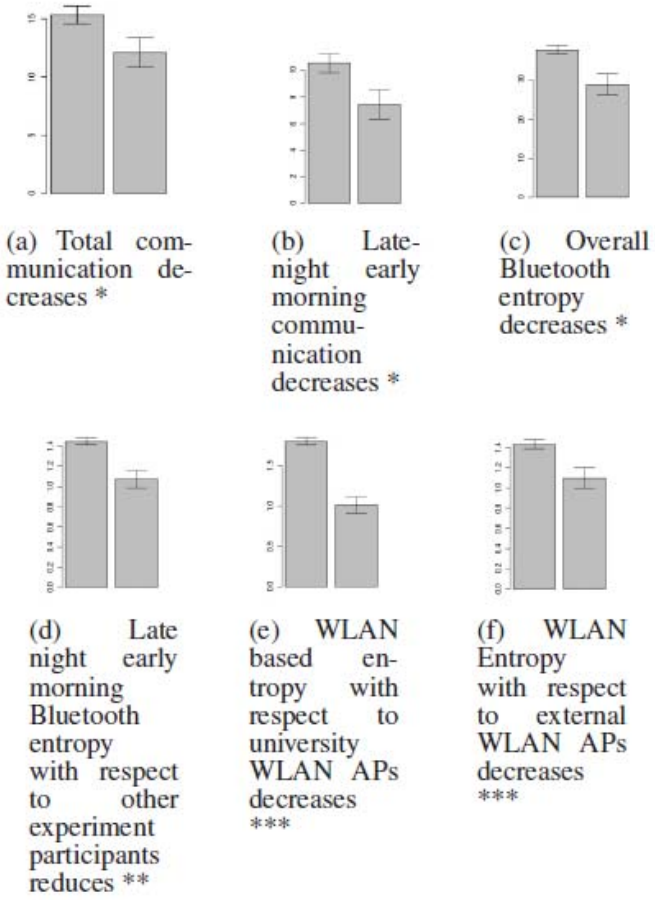
(c) WLAN based entropy with respect to university WLAN APs decreases ***

(d) WLAN Entropy with respect to external WLAN APs decreases ***

Figure 3. Behavior effects of fever, n=36/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

- Significant Drop in
  - Entropy of University WLAN APs
  - Entropy of external WLAN APs
- Moderate Drop in
  - Late Night/Early Morning Communications
  - Late Night/Early Morning Bluetooth entropy

Behavior Changes (Self-reported sad-lonely-depressed)



(a) Total communication decreases *

(b) Late-night early morning communication decreases *

(c) Overall Bluetooth entropy decreases *

(d) Late night early morning Bluetooth entropy with respect to other experiment participants reduces **

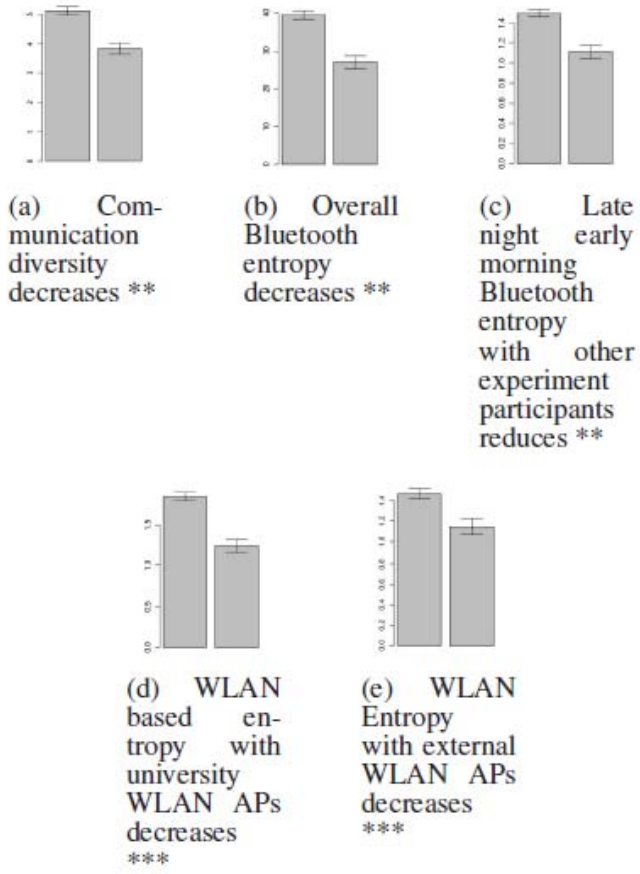
(e) WLAN based entropy with respect to university WLAN APs decreases ***

(f) WLAN Entropy with respect to external WLAN APs decreases ***

Figure 5. Behavior Changes with self-reported sad-lonely-depressed responses n=282/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

- Generally a decrease in mobile activity (isolation)

Behavior Changes (Self-reported often-stressed)



(a) Communication diversity decreases **

(b) Overall Bluetooth entropy decreases **

(c) Late night early morning Bluetooth entropy with other experiment participants reduces **

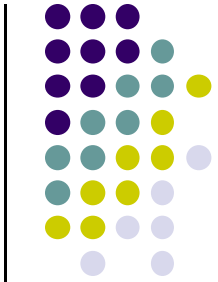
(d) WLAN based entropy with university WLAN APs decreases ***

(e) WLAN Entropy with external WLAN APs decreases ***

Figure 6. Behavior Changes with self-reported often-stressed responses n=559/2283, *: $p < 0.05$ **: $p < 0.01$ ***: $p < 0.001$

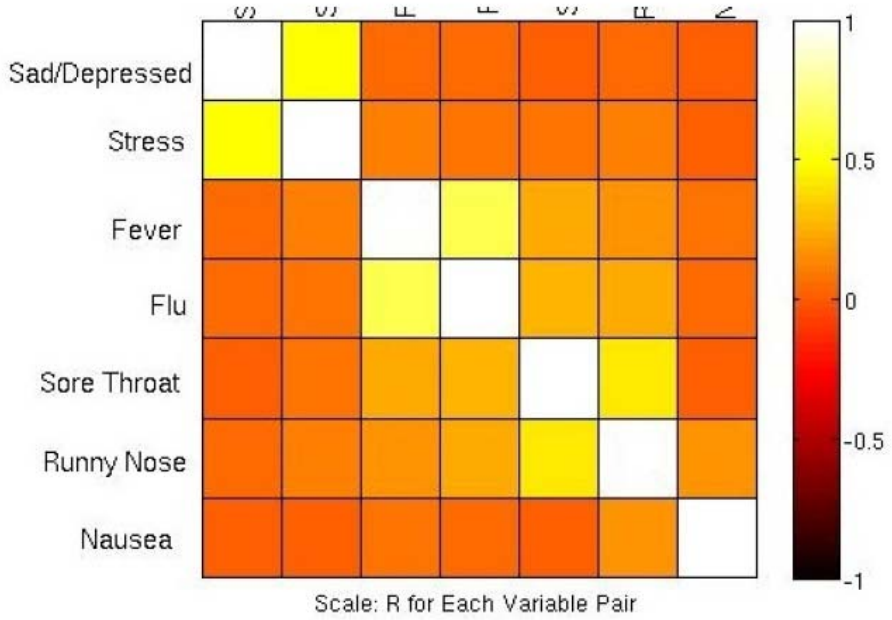
- Generally a decrease in mobile activity (isolation)

Symptom Classification



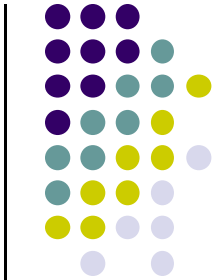
- Can we Build a Classification Scheme to Predict when Individuals are likely to be symptomatic
  - Asymmetric Misclassification Penalties (MetaCost)
    - Method for making classifiers cost-sensitive
  - Using K-Clustering, Four Clusters Emerge:
    - Stress + Depression
    - Runny Nose + Sore Throat
    - Fever + Influenza
    - Runny Nose + Sore Throat + Fever + Influenza

KNN Correlations Between Dependent Symptom Variables



- K-Nearest Neighbor
- Lighter Color Indicates Stronger Dependency
- Flu + Fever
- Runny Nose + Sore Throat
- Sad/Depressed + Stress

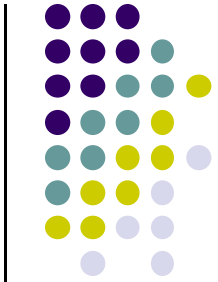
Statistical Terms



- Precision
  - Fraction of retrieved instances that are valid
- Recall
  - Fraction of relevant instances that are retrieved
- F-measure: combines precision and recall

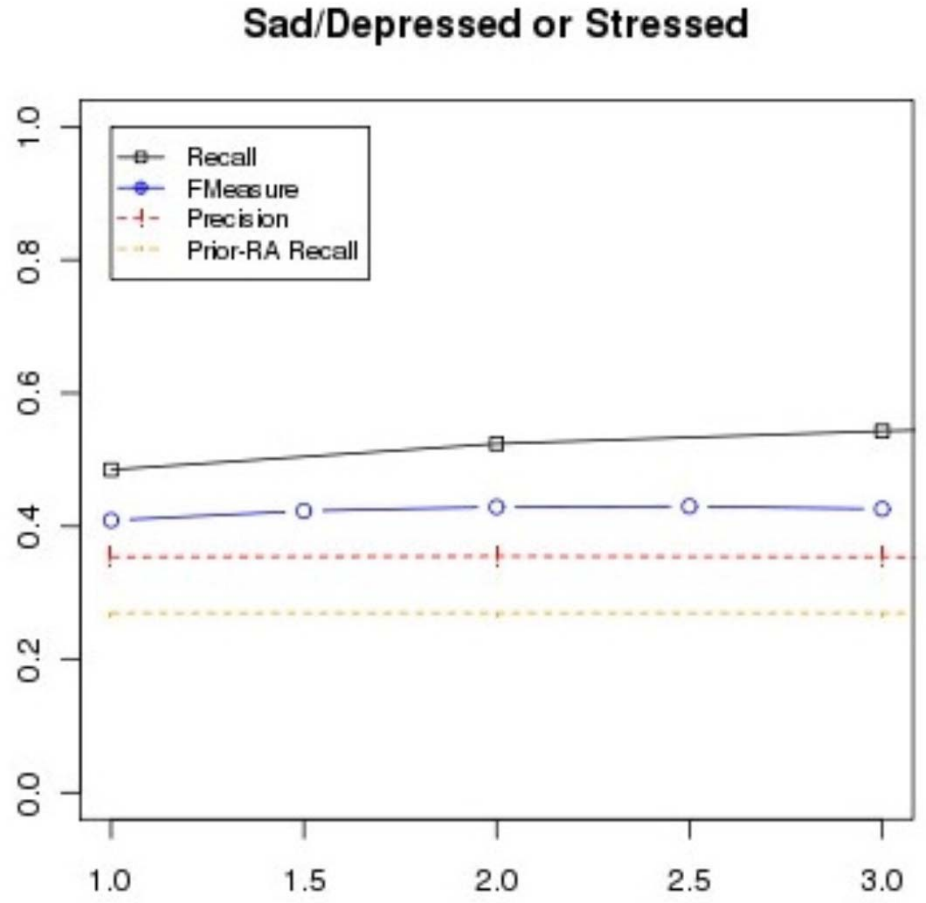
$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Example

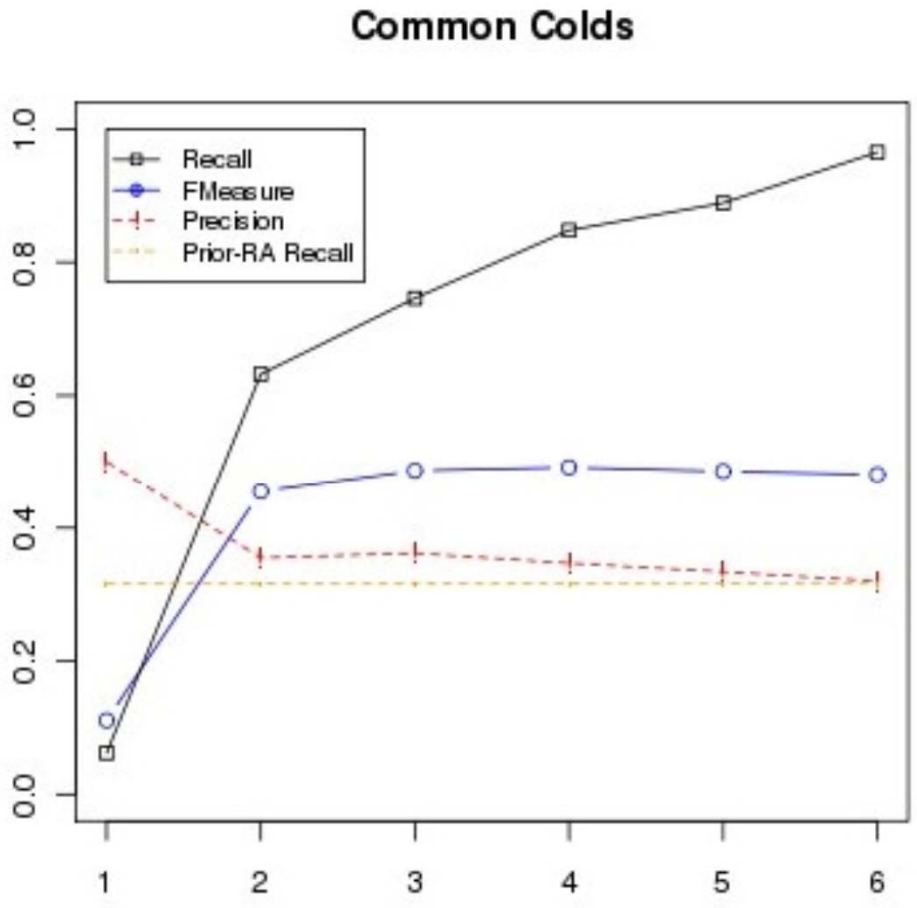


- When a search engine returns 30 pages only 20 of which were relevant while failing to return 40 additional relevant pages
  - Precision = 20/30 = 2/3
  - Recall = 20/60 = 1/3
  - F-measure = 2 * (2/3 * 1/3)/(2/3 + 1/3) = 4/9
- What does it mean when …
  - Precision is high but Recall is low?
  - Recall is high but Precision is low? Huh?

Classification Results

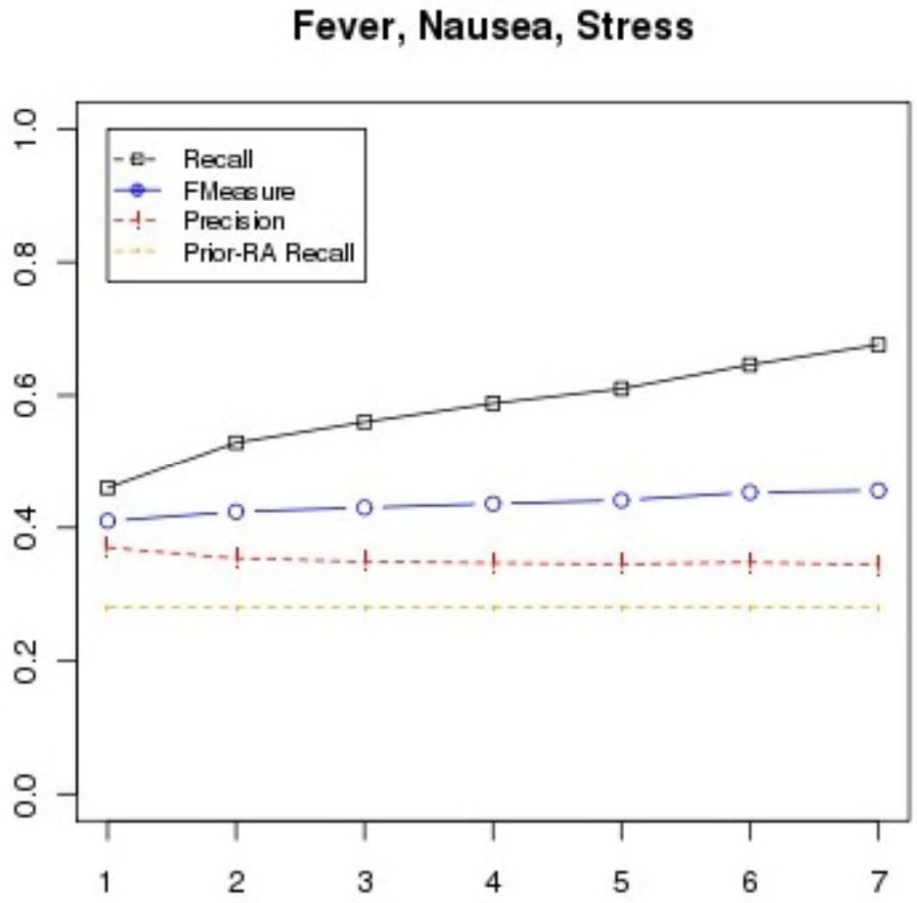


Classification Results

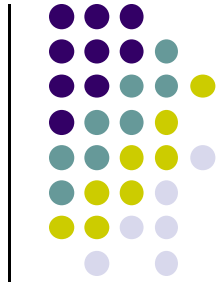


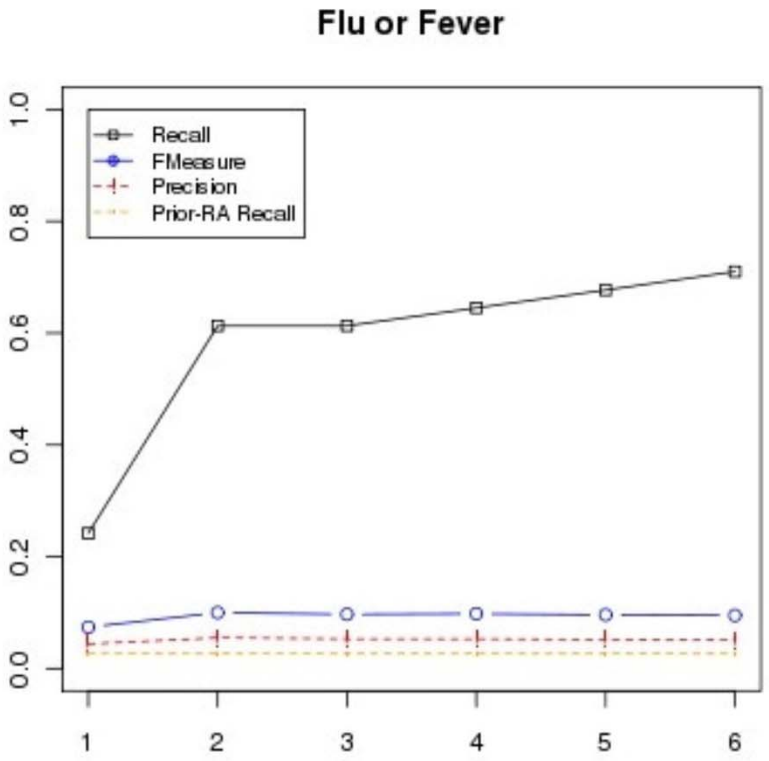
Sore-Throat, Cough,
Runny Nose, Congestion,
Sneezing Symptoms

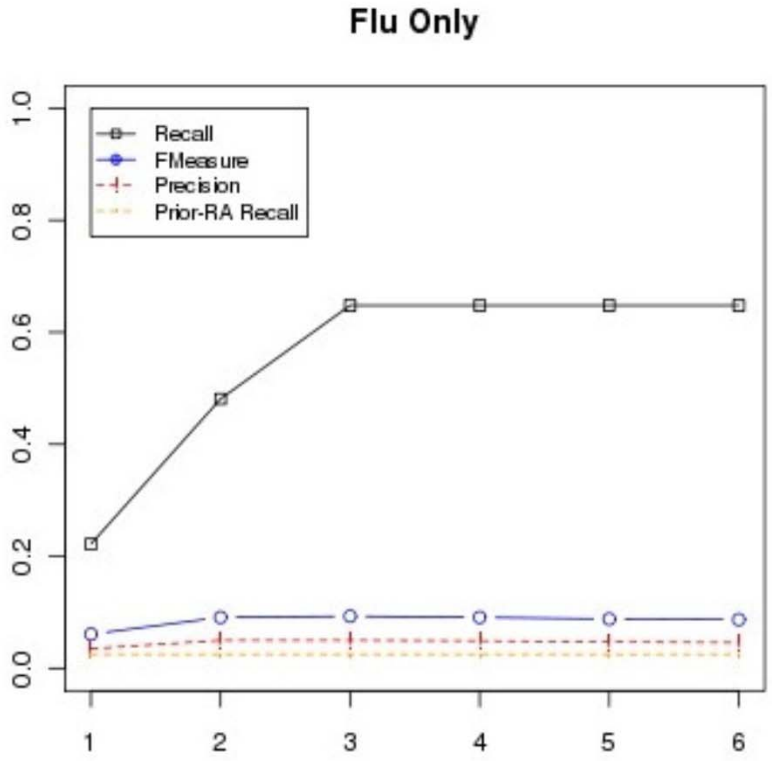
Classification Results



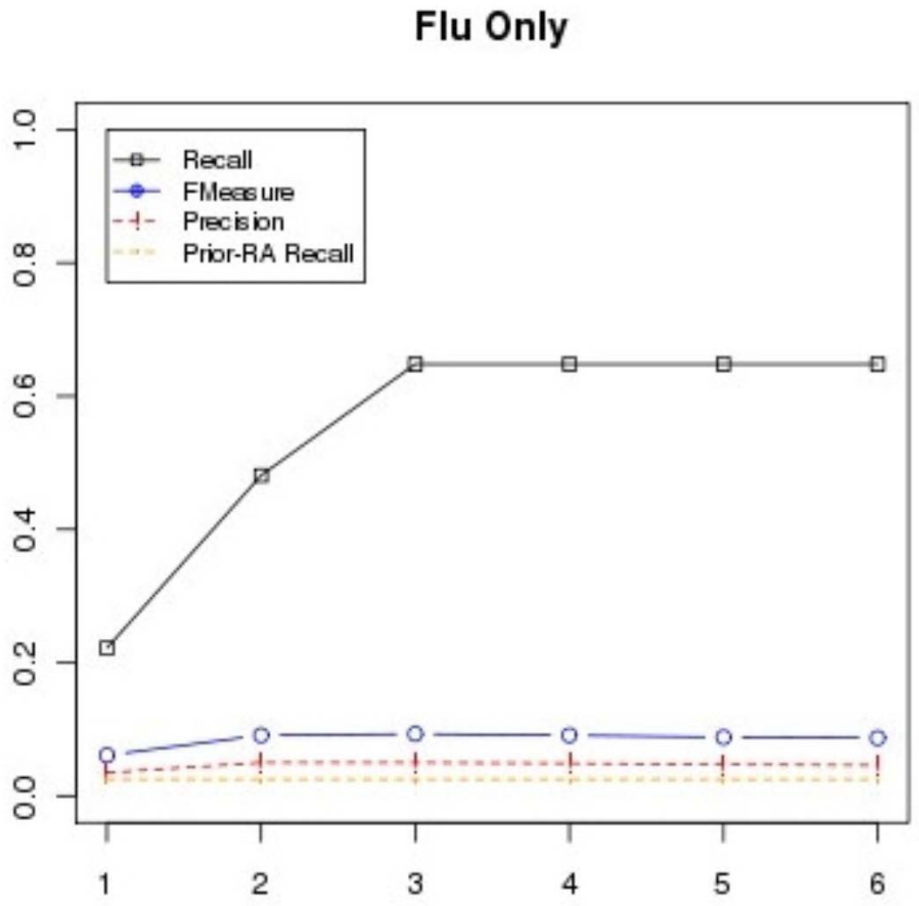
Classification Results



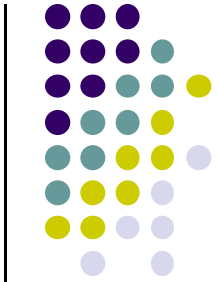




Classification Results

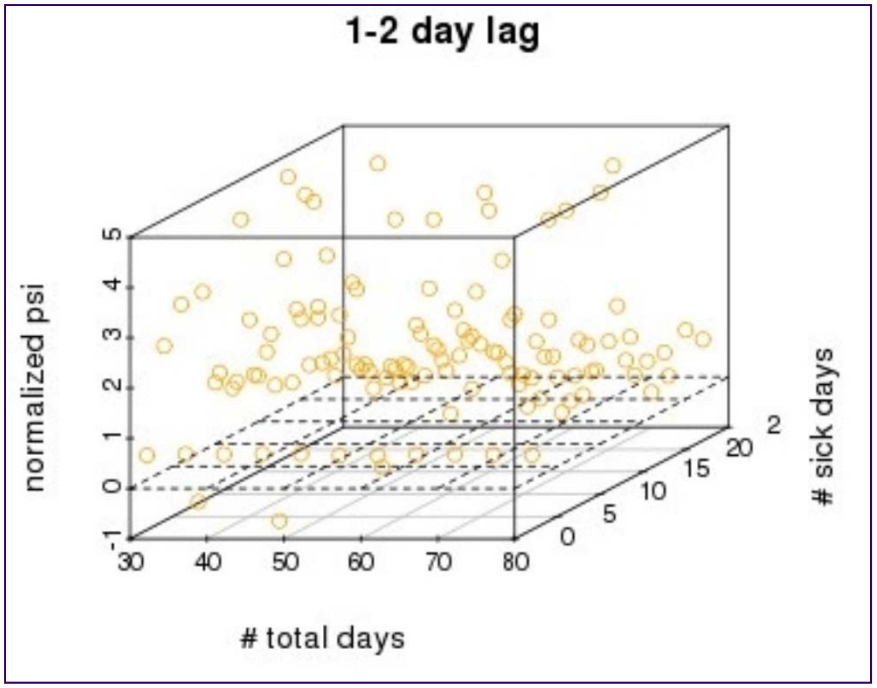


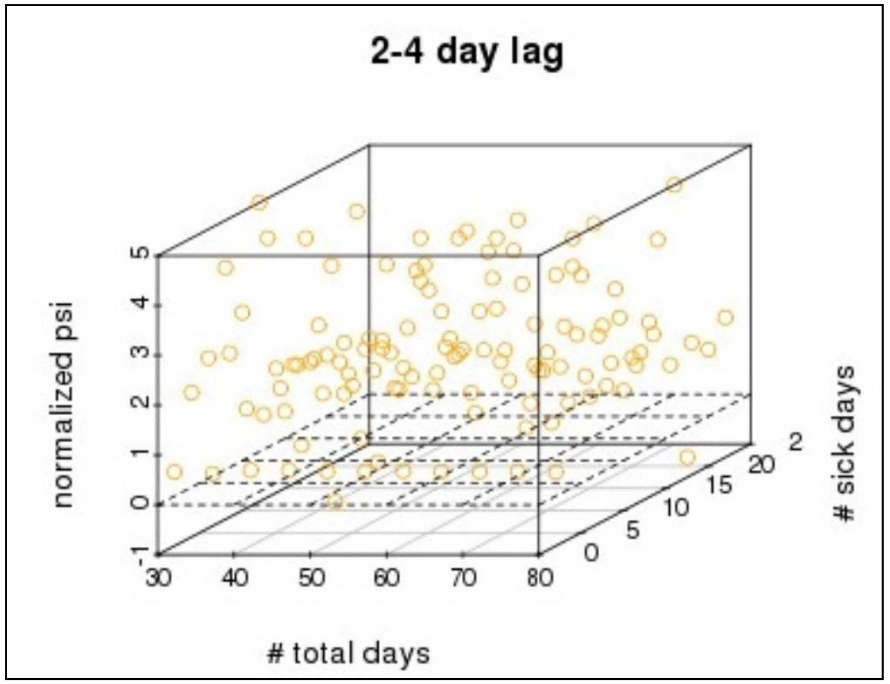
Granger Analysis & PSI



- Granger Analysis: determining whether one time series is useful in forecasting another
  - Predict a future event based on previous events
- Phase Slope Index (PSI) Method
  - More Noise Immune then Granger Analysis

PSI Evaluation on Lag

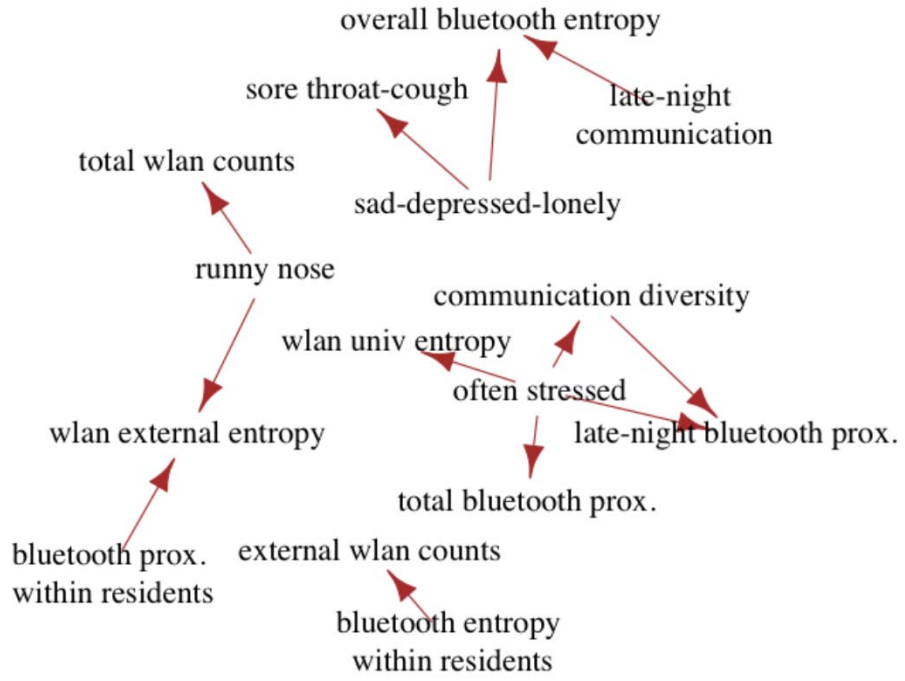




PSI Results

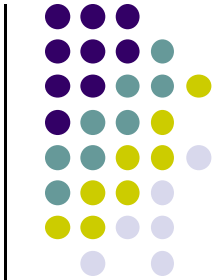
Table 2. PSI Results ordered by combined scores

Source	Follower
Runny nose	WLAN entropy with external APs
Sad-depressed-lonely	Sore throat-cough
Often stressed	Total Bluetooth proximity counts
Communication diversity	Late-night early morning Bluetooth proximity counts
Often stressed	Communication diversity
Often stressed	Late-night early morning Bluetooth proximity counts
Bluetooth entropy with other residents	External WLAN entropy
Runny nose	Total WLAN counts
Often stressed	WLAN entropy with university APs
Bluetooth proximity counts with other residents	External WLAN entropy
Late-night early morning communication	Overall Bluetooth entropy
Sad depressed lonely	Bluetooth entropy



'often-stressed' is useful in forecasting proximity, communication and WLAN behaviors, which suggests that individuals realize and report that they are stressed before it is reflected in their behavior

Conclusion

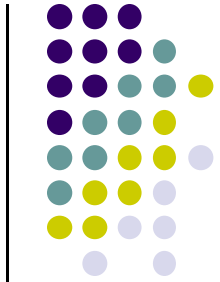


- Strengths
  - Shows the power of ubiquitous computing in Epidemiological Studies
  - K-Clustering and PSI Are Good Use of Predictive Models
  - Somewhat Dated, but the Idea is still relevant
    - Opens the door for further research (predictive healthcare)
- Weaknesses
  - Does not account for external factors like exams
  - Small Sample, homogenous Population (maybe)

References

- Madan, Anmol, et al. "Social sensing for epidemiological behavior change." *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010.
- Domingos, Pedro. "MetaCost: a general method for making classifiers cost-sensitive." *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999.
- *http://en.wikipedia.org/wiki/Epidemiology*
- *http://en.wikipedia.org/wiki/KNN*
- *http://en.wikipedia.org/wiki/Precision_and_recall*

Thank You



- Questions?