University of Nebraska at Omaha

Wireless Sensors and Big Data Analytics: A Focus on Health Monitoring and Civil Infrastructures

SENSORCOMM 2017

Hesham H. Ali
UNO Bioinformatics Core Facility
College of Information Science and Technology
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics and Heath Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
Tutorial Outlines

- **Scientific Data-Driven Revolution: An Overview**
- Big Data Analytics and Heath Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
Scientific Research will never be the same

- IT is changing many scientific disciplines
- So much relevant data is currently available
- The availability of data shifted many branches in sciences from pure experimental disciplines to knowledge based disciplines
- Incorporating Computational Sciences and other branches of sciences is not easy
- Interdisciplinary Research? Translational Research? Big Data Analytics?
The Future is Full of Opportunity

- Creating the future of networking
- Driving advances in all fields of science and engineering
- Revolutionizing transportation
- Personalized education
- The smart grid
- Predictive, preventive, personalized medicine
- Quantum computing
- Empowerment for the developing world
- Personalized health monitoring => quality of life
- Harnessing parallelism
- Synthetic biology
It’s all about the Data!

• How it all began:
  – Advances in instruments and computational technologies led to new research directions
  – Massive accumulation of data led to investigating new potential discoveries
  – The availability of enormous various types of public/private data sources
  – How to take advantage of the available data

• We are living the information world? Or the raw data world?
Challenges

• Too much data
• Are all available datasets relevant?
• Are they all accurate?
• Are they complete?
• Policy issues
• How to analyze such data
• Correlation versus causation
• How results can be verified? Validated?
Data Generation vs. Data Analysis/Integration

• New technologies lead to new data:
  – Competition to have the latest technology
  – Focus on storage needs to store yet more data

• Biomedical community needs to move from a total focus on data generation to a blended focus of measured data generation (to take advantage of new technologies) and data analysis/interpretation/visualization

• How do we leverage data? Integratable? Scalable?

• From Data to Information to Knowledge to Decision making
Current Focus on Data Generation
Data-Information-Knowledge-Wisdom

Increasing organisation
Increasing meaning(?)

Wisdom
- Applied Knowledge
  - Books, paradigms, systems, churches, philosophies
  - Schools of thought, poetry, belief systems, traditions, principles, truths
  - Weave, embody, discriminate, synthesize

Knowledge
- Organized Information
  - Chapters, theories, axioms, conceptual frameworks, complex stories, facts
  - Structure, interpret, evaluate, deconstruct

Information
- Linked Elements
  - Sentences, paragraphs, equations, concepts, ideas, questions, simple stories
  - Contextualise, compare, converse, connect, filter, prioritise, order, frame

Data
- Discrete Elements
  - Words, numbers, code, tables, databases
  - Categorise, calculate, collate, quantity, collect

Mapping

Design

Visualization

Smart Data Data-Driven Decisions

- **Data**: Physical entities at lowest abstraction level; contain little/no meaning – Measured data
- **Information**: Derived from data via interpretation – Processed data
- **Knowledge**: Obtained by inductive reasoning, typically through automated analysis and iterative collaboration – data + relationships
- **Decision Support**: 
A Focus on Biomedical Research

• Now we have data:
  – Advances in medical instruments and computational technologies led to new research directions
  – Massive accumulation of Biomedical data led to investigating new potential discoveries
  – The availability of enormous various types of public/private Biomedical data
  – How to take advantage of the available data

• Bioinformatics vs. Health Informatics vs. Biomedical Imaging vs. Public Health Informatics

• A new direction is now possible
Bioinformatics Data Sources
Mobility and Wellness

• Relationship between mobility and health is now well-documented
• Recent Advancement of Sensor technology
  Widespread of sensors
• Impact of Internet of Things
• So much relevant data is currently available
• Individual-focused mobility data
• The impact of the commercial aspect
• The potential of Big Data Analytics
Wellness Data Sources
Mobility and Health

• Mobility data collection: Continuous and non-invasive
• Mobility data: may not be 100% accurate
• Can we convert Individual-focused mobility data to population analysis
• From mobility parameters to health assessment to the prediction of health hazards
• Prevention and proactive healthcare as compared to reactive medicine
A Focus on Structural Health Monitoring (SHM)

• Process of determining and tracking structural integrity and assessing the nature of damage in a structure

• Approaches:
  – Assessment: Inspection (mostly by human inspection on yearly or biyearly basis)
  – Sensors Data collected but not used
  – Prediction: Very early stages

• SHM is a very complex “Big Data” problem
SHM: Current Challenges

• Numerous incidents of structural failures

• Bridges: Over 40K bridges in the US are not classified as safe!
  – Similar ratios in most countries
  – More data collection is taken place but not is not a critical part of the decision making cycle
Issues in Current Structural Health Monitoring

- Inaccurate & Incomplete
- Expensive & maintenance heavy
- Assessment accuracy increases with deterioration of infrastructures
- Experience with structural deterioration is not integrated in future assessment
A Focus on Bridge Monitoring

Increase safety, save costs & enhance efficiency of inspections

Focus on the core competences rather than administration

Make use of cutting-edge data mining technologies

Create a modern approach that incorporates sensor technology
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- **Big Data Analytics and Heath Monitoring**
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
What isn’t Necessarily Big Data?

• Just having more data (inevitable and boring!)
• A problem that can be solved by just more storage space
• Just producing or using large amounts of data
• Traditional schema-aware data and analysis
• Traditional 4-tier Architectures
“Big Data”

• **Big data:** “Any data too big to be handled by one computer” – Scientist John Rauser\(^1\)

• 90% of world's data created in last 2 years\(^2\)

• Data that reaches an order of magnitude that requires a new set of tools and methods to store, search, analyze, share, and visualize.\(^3\)

Updated Model

WISDOM

KNOWLEDGE

INFORMATION

DATA

NOISY DATA
The “Systems” Approach

- Integrated Approach:
  - Networks model relationships, not just elements
  - Discover groups of elements based on common relationships and/or properties

- Discovery
  - Examine changes in systems
    - Normal vs. diseased
    - Young vs. old
    - Fast vs slow
    - Stage I v. State II v. Stage III v. Stage IV
The Network “Graph” Model

- A network represents elements and their interactions
- Nodes $\rightarrow$ elements
- Edges $\rightarrow$ interactions
- Can represent multiple types of elements and interactions
Why Networks?

- Explosion of biological data

- Average microarray experiment: 1200 pages of data*

- How can we extract information from data?

* http://www.mc.vanderbilt.edu/peerreview/fall026.html
Critical Elements and Structures

Structure and function have been found to be related in social, biological, co-citation, transportation (etc.) networks\textsuperscript{1,2}.

The relationship between structure and function in the network model is unique to the construction of the model.

Population Analysis

- Able to handle ‘big’ data
- Draws from centuries of knowledge in graph theory
- Visually appealing and easy to understand
- When built correctly, structures can be tied to function
- Used in social, biological, technical applications
Health Monitoring - State of the Field

- Availability of many large useful devices – focus on collecting relevant data
- Availability of numerous helpful software packages
- Lack of data integration and trendiness of the discipline
- Fragmented efforts by computational scientists and biomedical scientists
- Lack of translational work – from the research domain to health care applications
- Increasing interest among researchers, industry and educators
Continuous and Comprehensive Monitoring

- **Dynamicity**
  - Physical activity level
  - Degree of activity

- **Variability**
  - Fluctuation in walking
  - Mobility Impairment

- **Symmetry**
  - Bilateral rhythm
  - Balance control functioning
“The future of biomedical research is about assimilating data across biological scales from molecules to populations. As such, the health of each one of us is a big data problem. Ensuring that we are getting the most out of the research data that we fund is a high priority for NIH.”

*Philip E. Bourne, Ph.D.*

*NIH associate director for data science*
Health of Individuals is a Big Data Problem

- Correlation graph using mobility parameters
Tutorial Outlines

• Scientific Data-Driven Revolution: An Overview
• Big Data Analytics and Health Monitoring
• *Wireless Sensors and Mobility Analysis for Healthcare*
• Correlation Analysis and Mobility – Network Analysis in Health Monitoring
• Civil Infrastructure and Data Analytics
• Technical Implementation Aspects of Network Analysis
• Next Steps – where to go from here?
Health and Mobility

• According to WHO: Health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.

• Mobility is broadly defined as the ability to move oneself
  – sedentary to high-intensity activities
  – local to high-level mobility
  – Without assistive devices to using assistive devices
How to collect mobility data

- Laboratory setting
- Real-world setting
- Self-reported data collection method
- Using monitoring devices, sensors and accelerometers
- GPS
- Gyroscope
Wearable monitoring and sensing devices

Main categories of wearable monitors:
• Pedometer
• Load transducer/foot-contact monitors
• accelerometers
• HR monitors
• Combined accelerometer and HR monitors
• Multiple sensor system
Pedometers versus Accelerometers

- Pedometers are less accurate in estimating distance and energy expenditure, but they are more accurate for step counting.
- For recording horizontal or vertical movement and measuring the frequency, duration, and intensity of physical activity, accelerometer is a better choice.
Tri-axial Accelerometer

- Differentiate gait patterns of older and younger adults while they were climbing stairs up and down.
- Identifying still postures, dynamic movement and postural transitions.
- Unable to determine the orientation of a body in the absence of movement.
- Integrating accelerometer with a gyroscope such as Shimmer.
Factors affecting health

- Age
- Gender
- Environmental factors
- Socioeconomic situation
- Family origin
- Physical Activity

*These factors are inter-related*
Physical Activities

• According to US Department of Health and Human Service, Physical activity all bodily movements produced by the contraction of skeletal muscle that increase energy expenditure above the basic levels.

• The first disease that was shown to be declined by regular physical activity was coronary heart disease.

• Physical activity reduces the risk of stroke, high blood pressure, type 2 diabetes, breast cancer, injurious falls, excessive weight gain, depression and loss of cognitive function.
Health outcomes associated with different Types of activity

<table>
<thead>
<tr>
<th>Examples of Physical activities</th>
<th>Examples of Physiologic changes</th>
<th>Examples of Health outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gardening</td>
<td>$\uparrow$ Autonomic balance</td>
<td>$\downarrow$ Breast cancer</td>
</tr>
<tr>
<td></td>
<td>$\uparrow$ Bone density</td>
<td>$\downarrow$ Colon cancer</td>
</tr>
<tr>
<td>Home repair</td>
<td>$\uparrow$ Capillary density</td>
<td>$\downarrow$ Coronary heart disease</td>
</tr>
<tr>
<td>Painting</td>
<td>$\uparrow$ Coronary artery size</td>
<td>$\downarrow$ Depression</td>
</tr>
<tr>
<td>Raking</td>
<td>$\uparrow$ Endothelial function</td>
<td>$\downarrow$ Excess weight gain</td>
</tr>
<tr>
<td>Shoveling</td>
<td>$\uparrow$ High density lipoprotein</td>
<td>$\downarrow$ Fractures</td>
</tr>
<tr>
<td>Sweeping</td>
<td>$\uparrow$ Immune function</td>
<td>$\downarrow$ Injurious falls</td>
</tr>
<tr>
<td>Vacuuming</td>
<td>$\uparrow$ Insulin sensitivity</td>
<td>$\downarrow$ Osteoporosis</td>
</tr>
<tr>
<td>Basketball</td>
<td>$\uparrow$ Lean body mass</td>
<td>$\downarrow$ Risk of death</td>
</tr>
<tr>
<td>Cycling</td>
<td>$\uparrow$ Mitochondrial volume</td>
<td>$\downarrow$ Stroke</td>
</tr>
<tr>
<td>Dancing</td>
<td>$\uparrow$ Motor unit recruitment</td>
<td>$\downarrow$ Type 2 diabetes</td>
</tr>
<tr>
<td>Running</td>
<td>$\uparrow$ Muscle fiber size</td>
<td>$\uparrow$ Cognitive function</td>
</tr>
<tr>
<td>Skiing</td>
<td>$\uparrow$ Neuromuscular coordination</td>
<td>$\uparrow$ Physical function</td>
</tr>
<tr>
<td>Soccer</td>
<td></td>
<td>$\uparrow$ Weight management</td>
</tr>
<tr>
<td>Swimming</td>
<td>$\uparrow$ Stroke volume</td>
<td></td>
</tr>
<tr>
<td>Tennis</td>
<td>$\downarrow$ Blood coagulation</td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>$\downarrow$ Inflammation</td>
<td></td>
</tr>
</tbody>
</table>
Classification of Physical Activities

• Quantitatively (frequency, volume and intensity)

• Qualitatively (sedentary, locomotion, work, leisure activities and exercise)

• Contextually (Posture, Time, Location, Social concepts - if you are doing an activity alone or with someone)
Wireless Sensors and Monitoring

- Human health can be significantly improved by monitoring the mobility patterns of individuals
- It is not easy to measure human activities because they are vary from person to person
- We developed a physical activity monitoring system using one wireless sensor
- Wireless sensor platform for this study is small, lightweight, and user-friendly
- We considered wireless communication and computation on this platform to minimize energy consumption
Wireless Networks in Aging

- Correlation between mobility and health level
- Monitoring mobility levels
- Aging of cells and aging of systems
- Collaboration between Bioinformatics group, Wireless Networks group and Decision Support Systems group
Falls and Associated Problems

- Falls are the leading cause of accidental deaths in the United States among people over the age of 75
  - the number of fatalities due to falls increased steadily from 14,900 in the year 2000 to 17,700 in 2005.

- Nebraska's over age 65 population is 13.3% versus 12.4% for the national average.
  - Generally speaking, the more rural the area, the higher the percentage of older adults.
  - In Nebraska, approximately 78% of those hospitalized for fall related injuries were 65 years and older.
Falling Problems

- Approximately 78% → 65 years and older.
- falls – leading cause of
  - injury deaths
  - injuries and trauma
- The risk of falling increases with age.
Wireless Sensor Based Mobility Monitoring

- Inexpensive
- Comfortable
- High mobility
- Simple

http://fiji.eecs.harvard.edu/Mercury
Four Phases of the Project

• Phase I: Fall Detection
  – achieved over 95% of fall detection rate

• Phase II: Classification of ADLs (Activities of Daily Living)
  – Running, Walking, Stair Climbing, Jumping, Stair Climbing, Standing, Sitting, and Lying, …

• Phase III: Construction of Mobility Profiles

• Phase IV: Fall (major health hazards) Prediction based on mobility profiles
Fall Detection Algorithm using Accelerometers
Mobility Sensors

- **Accelerometer**
  - Impact detection
  - (unit: gravity)

- **Gyroscope**
  - Measure rate of rotation
  - (unit: degrees per second)
Shimmers

- A wireless sensor platform for various types of wearable applications
- It consists of a number of integrated and extended sensors, a central processing unit, wireless communication module, and storage devices
- It has a low-power 8MHz MSP430 CPU, 10 KB RAM, 48 KB Flash memory, and 2 GB MicroSD card
- A 3-axis MMA 7361 accelerometer is integrated into Shimmer
Phase I: Fall Detection
Accelerometer-based fall detection

- Determine an acceleration threshold.
- Detect fall.
Phase I: Fall Detection

Accelerometer-based fall detection

- While the accelerometer-based algorithm is able to accurately detect major fall events, it also produced false positives for some events such as Jumping. Hence the need for learning components of the algorithm, or the use of additional sensors.
Phase II: Classification of ADLs

• Many Activities of Daily Living (ADLs) can be classified by analyzing the real-time acceleration data collected from sensors.

• Key Metrics
  – Inclination Angle
  – Standard deviation
  – Skewness
  – Signal Magnitude Area
Phase II: Classification of ADLs

• Many Activities of Daily Living (ADLs) can be classified by analyzing the real-time acceleration data collected from sensors.

• Data Processing: acceleration to metrics
  – Key Metrics
    • Inclination Angle
    • Standard deviation
Inclination Angle

- Inclination angle helps in determining posture.
  - Inclination angle is the measure between the x and y axis. It is assumed that if this value is around 180°, then the person would be standing.

- Can now differentiate standing, sitting, and lying.
Activity Classification

• Inclination angle has been selected to classify static activity like standing, sitting, and lying

• Calculates the angle with the x- and y-axis of accelerometer sensor

\[ \Phi = \arctan \frac{A_y}{A_x} \]

- \( A_x \): acceleration value of x-axis
- \( A_y \): acceleration value of y-axis
Inclination Angle

Angle Data

- Walking
- Standing
- Sitting
- Lying

Time
Standard Deviation

- Standard deviation of the x-axis acceleration helps in determining if the current mobility state is dynamic or static
  - Standard deviation measures the variability of data from the mean. Dynamic data will have measurably more variability than static data.

- When used with angle measurement, can now differentiate standing from walking/running.
Standard Deviation

- Standard deviation of the x-axis acceleration
  - $n$ is the number of points
  - $x_i$ is the acceleration at point $i$
  - $\bar{x}$ is the mean over the $n$ points

$$
\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
$$
Standard Deviation
Skewness

- Skewness helps in determining if the current mobility state is going up or down stairs.
  - Skewness measures the asymmetry of the distribution of x-axis acceleration values. It is assumed that going up/down stairs will produce data that has greater asymmetry than walking.
- When used with angle and standard deviation, can now differentiate walking/running from going up/down stairs.
Skewness

- Skewness of the x-axis
  - $n$ is the number of points
  - $x_i$ is the acceleration at point $i$
  - $\bar{x}$ is the mean over the $n$ points
  - $\sigma$ is the standard deviation over the $n$ points

$$skewness = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$$
Skewness

Signal Magnitude Area

- Measures amplitude and duration variation in the acceleration signal.
  - Assumed that amplitude and duration variation will be greater when the intensity of the activity changes. As such, SMA values will be greater when changing state as opposed to not changing state. Ex: getting out of bed vs. walking

- When used with the prior four measurements, SMA will help differentiate transitions from other dynamic mobility states.
Signal Magnitude Area

- SMA is the sum of the integrals of the three axis over time period T.
  - $a_x$, $a_y$, and $a_z$ are the accelerations of the x, y, and z axis.
  - normalized by dividing by length $t$.

\[
SMA = \frac{1}{t} \left( \int_{t=0}^{T} |a_x| \, dt + \int_{t=0}^{T} |a_y| \, dt + \int_{t=0}^{T} |a_z| \, dt \right)
\]
Phase III: Mobility Profiles - Signatures

- Is it possible to develop a mobility signature to represent the makeup of someone’s mobility over a given period of time?
- Can the profile be tracked over longer periods to assess progress or lack thereof in terms of mobility – or health levels?
- Is it possible to compare the mobility of health level of individual as compared to a given populations?
## Experiment Schedule

<table>
<thead>
<tr>
<th>Activity</th>
<th>Place</th>
<th>Time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subject I</td>
</tr>
<tr>
<td>Walking</td>
<td>Outdoor</td>
<td>360</td>
</tr>
<tr>
<td>Standing</td>
<td>Outdoor</td>
<td>20</td>
</tr>
<tr>
<td>Running</td>
<td>Outdoor</td>
<td>30</td>
</tr>
<tr>
<td>Standing</td>
<td>Outdoor</td>
<td>30</td>
</tr>
<tr>
<td>Walking</td>
<td>Outdoor</td>
<td>300</td>
</tr>
<tr>
<td>Sitting</td>
<td>Indoor</td>
<td>360</td>
</tr>
<tr>
<td>Walking</td>
<td>Indoor</td>
<td>30</td>
</tr>
<tr>
<td>Lying</td>
<td>Indoor</td>
<td>72</td>
</tr>
<tr>
<td>Total time</td>
<td></td>
<td>1200</td>
</tr>
</tbody>
</table>
Profiles: Experiment Results

**Test**
- Walking: 54%
- Sitting: 32%
- Standing: 5%
- Lying: 7%
- Running: 2%

**Goal**
- Walking: 57%
- Sitting: 30%
- Standing: 4%
- Lying: 6%
- Running: 3%
Phase IV: Fall Prediction

- Fall can injure the elderly in large scale.
- 10-15% falls cause some serious physical injury in older people.
- The early prediction of fall is an important step to alert and protect the subject to avoid injury.
- We employ Hidden Markov Models for detection and prediction of anomalous movement patterns among the human subjects.
Phase IV: Fall Prediction

User's Mobility Profiles

Fall-Prediction Module (Hidden Markov model)

Training Mobility Profiles

User's Mobility Health Index
Prediction Model

• Develop a HMM model with
  – 3 states (Healthy, Unhealthy, Mildly Unhealthy)
  – 3 parameters (#steps moved, #rooms visited, # movement in arms)

• Entire dataset was split into
  – Train data
  – Test data
THREE-STATE TRANSITION DIAGRAM
Summary

• Proposed an on-board processing approach for classifying Activity of Daily Living using a triaxial acceleration sensor
• Implemented this mechanism on a tiny wireless sensor which is easy to wear and user-friendly
• Integrated all core functionalities such as an activity classification, wireless communication module, and data storage function into a single wireless sensor platform
• Signatures for disease or recovery from operations
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics: An Overview
- Heath Monitoring and Data Analytics
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Networks and Population Analysis
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics and Health Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- *Correlation Analysis and Mobility – Network Analysis in Health Monitoring*
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
Using Correlation Networks

- Correlation graph using mobility parameters
• **Correlation networks**
  – What are they? How are they made?
  – How can they be used in biomedical research?

• **Network Comparison**
  – Identifying common & unique network elements
  – Filtering noise from causative relationships

• **Case study**
  – Proof of concept using sample expression data
  – What kinds of questions can we ask?
Power of Correlation Analysis

- Correlation versus Causation
- Correlation networks and population analysis
- Casting the net wide – signal and noise
- The use of enrichment before obtaining information and after for validation
Correlation Network Applications

• “Versus” analysis
  • Normal vs. disease
  • Times/environments

• Model for high-throughput data
  • Especially useful in microarrays

• Identification of groups of causative genes
  • Ability to rank based on graph structure
  • Identify sets of co-regulated, co-expressed genes
Correlation Network Analysis

Correlation = -1

Correlation = 0

Correlation = 1

Gene 1

Gene 2

Gene 1

Gene 2

Sample

Expression Level

-0.99

0.00

0.99
Local Network Structures

• Cliques
  Protein complexes, regulatory modules

• Pathways
  Signaling cascades

• Hubs
  Regulators, TFs, active proteins

• Articulation points – Gateways
Network Concepts

• Biological networks have structural properties
  • Can differ from one network to another

• Specific structures/characteristics have biological meaning
  • Degree can indicate essentiality
  • Cluster density can indicate relevance

• Networks do not have to be static
  • Most interesting discoveries coming from temporal or state-change network alignment & comparison
Hypothesis

Correlation networks are an excellent tool for mining relationship rich knowledge from high-throughput data

Using systems biology approach, CN can help identify:
- Critical Genes that are essential for survival
- Subsets of genes that are responsible for biological functions

Measures of centrality to identify key elements:
Proves existence of structure/function relationship in correlation networks
Big Data Analytics

• The novelty of our system lies not in the technology itself (i.e. the sensor), but rather in the data analysis

  – Correlation network for finding clusters of highly correlated attributes of Gait Signature Metrics to describe the correlation patterns among different health conditions
Mobility Analysis

- Human Mobility
  - Clinical significance
  - Related to various aspects of health

- Characteristics of Human Mobility
  - Natural variability
  - Non-deterministic nature

- Population based analytics
  - Identifying mobility patterns changes
  - e.g. preventive treatment for emerging health hazards
Health Assessment Using Activity Classification

• Mobility parameter derived using the weighted average activities performed and the metabolic (or a unified parameter such as Energy Expenditure) equivalents of these activities.

• Weighted average of the time for which a particular activity is performed is captured, and the metabolic equivalent is used to assign the point to gauge mobility.

• The importance of Context
The edge in a network represents the mobility parameter. It can be one of the following:

- Step Length
- Step Time
- Number of steps
- Variability
- Dynamicity
- Symmetry
- Energy dissipated
Mobility Parameters Using Activity Classification

Daily Activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting</td>
<td>26%</td>
</tr>
<tr>
<td>Standing</td>
<td>25%</td>
</tr>
<tr>
<td>Walking</td>
<td>13%</td>
</tr>
<tr>
<td>Jumping</td>
<td>13%</td>
</tr>
<tr>
<td>Running</td>
<td>8%</td>
</tr>
<tr>
<td>Downstairs</td>
<td>4%</td>
</tr>
<tr>
<td>Upstairs</td>
<td>13%</td>
</tr>
<tr>
<td>Sleeping</td>
<td>13%</td>
</tr>
</tbody>
</table>

Physical Activity

<table>
<thead>
<tr>
<th>Activity</th>
<th>METS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking (2 mph)</td>
<td>2.5</td>
</tr>
<tr>
<td>Dancing</td>
<td>2.9</td>
</tr>
<tr>
<td>Walking (3 mph)</td>
<td>3.3</td>
</tr>
<tr>
<td>Walking (4 mph)</td>
<td>4.5</td>
</tr>
<tr>
<td>Jogging (10 min miles)</td>
<td>10.2</td>
</tr>
<tr>
<td>Climbing</td>
<td>6.9</td>
</tr>
<tr>
<td>Lying</td>
<td>1</td>
</tr>
<tr>
<td>Sitting</td>
<td>1</td>
</tr>
<tr>
<td>Standing</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Case Study

- Simulation Study
  - Mobility of nurses in a hospital – 8 hour shifts versus 12 hour shifts
  - Monitoring mobility pattern changes at different times during the shift
- Experimental Study
  - Mobility of mice in a cage
  - Identifying/classification of various groups based on mobility characteristics
### Case Study data

- **Sample Generation Setting**
  - Weighted activity level value
  - Each group has different mobility decline rate per hour
  - Group 1 - 10%/hour, group 2 - 20%/hour and group 3 - 30%/hour (shown in different colors)

<table>
<thead>
<tr>
<th></th>
<th>Sub1</th>
<th></th>
<th></th>
<th>Sub30</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work start</strong></td>
<td>553.78</td>
<td>-</td>
<td>-</td>
<td>384.85</td>
</tr>
<tr>
<td><strong>2nd hours</strong></td>
<td>498.40</td>
<td>-</td>
<td>-</td>
<td>269.40</td>
</tr>
<tr>
<td><strong>4th hours</strong></td>
<td>448.56</td>
<td>-</td>
<td>-</td>
<td>188.58</td>
</tr>
<tr>
<td><strong>6th hours</strong></td>
<td>403.71</td>
<td>-</td>
<td>-</td>
<td>132.01</td>
</tr>
<tr>
<td><strong>Work end</strong></td>
<td>363.34</td>
<td>-</td>
<td>-</td>
<td>92.40</td>
</tr>
</tbody>
</table>
Scenario Description

• Analyzing clusters from correlation networks
• Networks are constructed for every mobility samples captured from nurses at 4 different times as the day progresses.
• Magnitude based analysis applied
• Different levels of mobility decline in nurses identified from networks.
Correlation Networks

Sample 1

Green nodes – Group 1
Pink nodes – Group 2
Red nodes – Group 3

Sample 2
Time Series

Sample 3

Green nodes – Group 1
Pink nodes – Group 2
Red nodes – Group 3

Sample 4
Simulation Results

• As the time progressed in the day, the same colored nodes (nurses) formed a separate clusters according to their raw mobilities.
• This network shows different clusters each of low (red), medium (pink) and high (green) mobilities.
• Identified different mobility features
• Application to predict medical hazards
• Practical to free-living environment
Health Assessment using Gait Parameters

- Gait parameters and health conditions
- Traditional gait measurement systems
- Advanced technologies
- Continuously and remotely monitoring
Goals and Applications

• **Main Goal**
  – Assess health levels using gait parameter

• **Applications**
  – Diagnosis of diseases in the early stages
  – Monitoring elderlies
  – Rehabilitation
  – Evaluation of treatment options

• **Gap**
  – Lack of a rigorous model that links health levels to the mobility patterns
• **Main Question**
How to utilize individuals’ gait patterns and mobility profiles to assess health levels?

• **Approach**

✓ Population Analysis
✓ Similarity measure
  - Gait parameters
  - Age
  - Gender
  - Genetic backgrounds
✓ Network Model

- Identifying and Extracting Discriminating Parameters
- Performing Similarity Analysis and Creating Network Model
Identifying Health Conditions Affecting Gait Patterns

- Parkinson’s Disease
- Multiple Sclerosis
- Amyotrophic Lateral Sclerosis
- Huntington’s disease
- Aging
- ...
- Healthy Control, Geriatrics, patients with PD
Identifying Discriminating Gait Parameters

- Gait Parameters Discriminating between Healthy Younger individuals and Geriatrics
  - Shorter step length
  - Increased double support time
  - Less power at ankle
  - Less pelvic tilt and hip flexion-extension
  - Stride time variability
  - Asymmetry of the acceleration in the AP direction
Identifying Discriminating Gait Parameters

- Gait Parameters Discriminating between Healthy Younger individuals and Patients with PD
  - Longer double limb support
  - Shorter stride length
  - Greater stride time
  - Increased left-right gait asymmetry
  - Higher gait variability
Participants and Protocol

- **Participants**
  
<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>PD</th>
<th>Geriatrics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of subjects</strong></td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Gender (M/F)</strong></td>
<td>3:2</td>
<td>3:2</td>
<td>2:3</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>64 (10)</td>
<td>72 (6.3)</td>
<td>81 (5.9)</td>
</tr>
<tr>
<td><strong>UPDRS III</strong></td>
<td></td>
<td>20.8 (6.1)</td>
<td></td>
</tr>
<tr>
<td><strong>H &amp; Y</strong></td>
<td></td>
<td>2.6 (0.5)</td>
<td></td>
</tr>
</tbody>
</table>

- **Protocol**
  - 2 minutes free walk around the hospital (two times)
  - SHIMMER (left and right ankles)
  - 40 seconds
Left and right ankle Acceleration (AP) of three subjects from three groups
Variability, Symmetry and Intensity

• Variability
  – Standard deviation
  – Coefficient of variation

• Intensity

\[
\text{Intensity} = \sum x^2 + y^2 + z^2
\]

• Symmetry

\[
\text{Symmetry} = \frac{\text{Min}(Fl, Fr)}{\text{Max}(Fl, Fr)}
\]
## Stride Segmentation and Parameter Extraction

<table>
<thead>
<tr>
<th>Parameter</th>
<th>What this parameter measures</th>
<th>Axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg_stride_Time</td>
<td>Average stride Time</td>
<td>-</td>
</tr>
<tr>
<td>Acc_Var</td>
<td>Standard deviation of stride_to_stirde acceleration (stride-to-stride acceleration <strong>variability</strong> )</td>
<td>X, Y and Z</td>
</tr>
<tr>
<td>CV_StrideTime</td>
<td>Stride time coefficient of <strong>variation</strong></td>
<td>-</td>
</tr>
<tr>
<td>Avg_RMS</td>
<td>Magnitude of acceleration (<strong>Intensity</strong>)</td>
<td>X, Y and Z</td>
</tr>
<tr>
<td>Max_Acc_perStride</td>
<td>Mean value of maximum magnitude of acceleration per stride</td>
<td>X, Y and Z</td>
</tr>
<tr>
<td>Avg_VecMag</td>
<td>Vector magnitude of acceleration in all directions (<strong>Intensity</strong>)</td>
<td>-</td>
</tr>
<tr>
<td>Stride_time_sym</td>
<td>Stride time <strong>symmetry</strong> between left and right leg</td>
<td>-</td>
</tr>
<tr>
<td>VecMag_sym</td>
<td>vector magnitude <strong>symmetry</strong> between left and right leg</td>
<td>-</td>
</tr>
</tbody>
</table>
### ANOVA and Post Hoc test

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Control vs Geriatrics</th>
<th>Control vs PD</th>
<th>Geriatrics vs PD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference</td>
<td>sig.</td>
<td>Mean difference</td>
</tr>
<tr>
<td>Avg_stride_time</td>
<td><strong>-0.23</strong></td>
<td><strong>0.004</strong></td>
<td>0.53</td>
</tr>
<tr>
<td>Var_Acc_X</td>
<td>0.037</td>
<td>0.075</td>
<td>0.41</td>
</tr>
<tr>
<td>Var_Acc_Y</td>
<td>0.049</td>
<td>0.063</td>
<td>-0.19</td>
</tr>
<tr>
<td>Var_Acc_Z</td>
<td>0.077</td>
<td>0.051</td>
<td>0.02</td>
</tr>
<tr>
<td>CV_StrideTime</td>
<td><strong>-0.22</strong></td>
<td><strong>0.010</strong></td>
<td><strong>-0.28</strong></td>
</tr>
<tr>
<td>Avg_VecMag</td>
<td><strong>0.04</strong></td>
<td><strong>0.004</strong></td>
<td>-0.21</td>
</tr>
<tr>
<td>Avg_RMSX</td>
<td><strong>0.37</strong></td>
<td><strong>0.000</strong></td>
<td>0.039</td>
</tr>
<tr>
<td>Avg_RMSY</td>
<td><strong>0.43</strong></td>
<td><strong>0.001</strong></td>
<td>-0.14</td>
</tr>
<tr>
<td>Avg_RMSZ</td>
<td><strong>0.19</strong></td>
<td><strong>0.040</strong></td>
<td>0.33</td>
</tr>
<tr>
<td>Avg_MaxAcc_X</td>
<td><strong>1.67</strong></td>
<td><strong>0.000</strong></td>
<td><strong>0.72</strong></td>
</tr>
<tr>
<td>Avg_MaxAcc_Y</td>
<td><strong>1.87</strong></td>
<td><strong>0.000</strong></td>
<td>0.38</td>
</tr>
<tr>
<td>Avg_MaxAcc_Z</td>
<td>1.01</td>
<td>0.140</td>
<td>-0.46</td>
</tr>
<tr>
<td>Sym_VecMag</td>
<td>0.42</td>
<td>0.323</td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>Sym_StrideTime</td>
<td>0.32</td>
<td>0.080</td>
<td>0.28</td>
</tr>
</tbody>
</table>
Correlation Analysis and The Network Model

- **Correlation Analysis**
  - A pairwise Pearson correlation analysis between subjects, using gait parameters
  - Threshold $\rightarrow$ 95.5 %
  - Significance $\rightarrow$ 0.05

- **Creating Network Model**
  - Vertices represent subjects
  - If two subjects are highly correlated, there is an edge
The Network Model

- Clusters
- Cliques
- Gateway Vertices
- Transitions
The Dynamic Aspect
Summary

- This model worked well when we applied it to our dataset
- Two clusters (healthy vs others)
- No clear distinction between geriatrics and PD
  - We did not pick gait parameters discriminating between these two groups
  - Their gait parameters are pretty similar
- Gateway node 6 → might be due a transition between health conditions!! (healthy to geriatrics)
- Limitation → 15 subjects
Back to Network Analysis

- Correlation graph using mobility parameters
Applications in hospital: Post-operative Nursing Care

- A post-operative assessment is very important to a full and speedy recovery from any type of surgery.
  - a full assessment and an individualized treatment plan based upon the patient’s needs and level of function, coupled with clinician expectations
Applications for Patient Care: Patient care after discharge

• By analyzing mobility profiles created by the proposed device, we would be able to
  – measure changes in daily/hourly activity levels,
  – identify anomalous movement and patterns,
  – predict and possibly prevent future mobility problems
Tutorial Outlines

• Scientific Data-Driven Revolution: An Overview
• Big Data Analytics and Heath Monitoring
• Wireless Sensors and Mobility Analysis for Healthcare
• Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
• Civil Infrastructure and Data Analytics
• Technical Implementation Aspects of Network Analysis
• Next Steps – where to go from here?
A Focus on Bridge Monitoring

Increase safety, save costs & enhance efficiency of inspections

Focus on the core competences rather than administration

Make use of cutting-edge data mining technologies

Create a modern approach that incorporates sensor technology

Research Goals
A Focus on Structural Health Monitoring (SHM)

• Process of determining and tracking structural integrity and assessing the nature of damage in a structure

• Approaches:
  – Assessment: Inspection (mostly by human inspection on yearly or biyearly basis)
  – Sensors Data collected but not used
  – Prediction: Very early stages

• SHM is a very complex “Big Data” problem
SHM: Current Challenges

• Numerous incidents of structural failures

• Bridges: Over 40K bridges in the US are not classified as safe!
  – Similar ratios in most countries
  – More data collection is taken place but not is not a critical part of the decision making cycle
Issues in Structural Health Monitoring

• Inaccurate & Incomplete
• Expensive & maintenance heavy
• Assessment accuracy increases with deterioration of infrastructures
• Experience with structural deterioration is not integrated in future assessment
# Health Monitoring through Assessment

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Assessed through</th>
</tr>
</thead>
</table>
| Global    | + Assessment of overall Structure Health
- Often does not detect damage
- Vulnerable to environmental effects
Examples Measures: Structural Mode Shapes | Matrix Update Method
Image Processing
Ultrasonic Measurement
Impact-Echo
Tap Tests
Reduction of Noise Through
• Baseline Signal Reduction
• Hilbert-Huang Transformation
• Actuators and Sensors |
| Local     | + More Precise
- Too Costly Effort | Acoustic Waves
X-Rays
Radar |
Anomaly detection with graphs

- Interactive graph querying:
  - Left: Given a set of detected abnormal nodes
  - Right: Further grouping into few clusters. Detecting how the abnormal nodes are linked to each allows to find undetected abnormalities with similar characteristics
Anomaly Detection Example 1: Monitored Life Load Effects

- Introduces an approach to evaluating safety of existing bridges through structural health monitoring (SHM)
  - **Condition function**
    estimates strains at locations other than the strain gauge locations
  - **Prediction function**
    predict extreme values of the SHM data in future
  - A single sensor can measure traffic, load, weight, speed of traffic,…
Anomaly Detection Example 2: Railway Track Performance Monitoring

• Addition to previous contribution by thermal stress

• Provides ideas on how to implement / integrate
  – temperature sensors / data
  – Analysis of temperature changes
  – Safety Warning System
  – Temperature alarm values

• Relevance for this research
  – Bridge Membrane Monitoring
  – Steel Constructions
Bridge Correlation Networks

Raw Data → Correlation Network → Cluster Analysis
The Graph Model

**Nodes** represent **bridges** and can be styled with bridge data. E.g. size, color, shape…

**Edges** represent **similarities** or shared characteristics between bridges and can be styled as well.
Key Research Questions

• What are the main features that impact structural safety?
• What is the best granularity level to conduct the safety analysis?
• How to use available data (historic information) to make better decisions?
• How to improve structural safety and performance using big data analytics?
National Bridges Inventory

Recording
Archiving

Policies
Information
Guidance

Inspections
& Costs

Official NBI Data
http://www.fhwa.dot.gov/bridge/nbi.cfm

Online Database (NBI & Bridgehunter)
http://uglybridges.com
Anomaly Detection with Graphs

- Interactive graph querying:
  - Left: Given a set of detected abnormal nodes
  - Right: Further grouping into few clusters. Detecting how the abnormal nodes are linked to each allows to find undetected abnormalities with similar characteristics
Dynamic Graph Models

- Sensor inputs
- Additional data sources
- Data manipulation

Clustering / Prediction Model

Generic Approach
- no data preparation
- Can find unknown problems

Specific Approach
- faster learning
- more in-depth analysis

\[ \text{sim}_{a,b} = \sum_{i=1}^{n} c_i \cdot v_i \]

\[ \Rightarrow \text{Minimize} \]
\[ \Rightarrow \text{Maximize} \]
Visualization Age / Safety

SR = Sufficiency Rating
AGE = Bridge Age (or last reconstruction)

Correlations based on:
- Age
- SR

Coloring based on:
- SR
Prediction: Genetic Algorithms

- Initial Population
- Reproduction
- Determination of Fitness
- Results
- Mutation
- Selection
Nebraska Bridges 2011-2014

N= random, 1000 bridges
(46% visible, filtered unconnected nodes)

Conditions:
• Delta SR between 2 and 6
• Delta Deck Condition > 3

Colored by SR 2014

Correlations based on:

Coloring based on:

SR
CON
SR 14
Summary of Results

• Early study confirms the impact of key parameters:
  – Age
  – Recent reinforcements/maintenance
  – Traffic
• Other critical parameters:
  – Type/technology of bridges
  – Variability in temperatures – extreme temps
  – Rate of change – Delta analysis
• Population analysis leads to better inspection or maintenance schedule
• Big data analysis is critical to the development of smart structures
Next Steps

• Include more parameters in the study
  – A focus on temperature variability

• Conduct the analysis at different granularity levels
  – A focus on structural cracks

• Build a decision support system
  – Collaborate with state transportation experts

• Explore the utilization of advanced graph theoretical analysis
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics and Heath Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- *Technical Implementation Aspects of Network Analysis*
- Next Steps – where to go from here?
How to implement this stuff?

Computational Issues

- Graph/Network Modeling
- Graph Algorithms
- High Performance Computing
  - Beyond surface-level adaptation of known algorithms
- Wireless Networks
- Statistical Analysis
- Storage/processing models - Security and Privacy
HPC and Big Data

• Network creation: 2 weeks on PC
  • 10 hours in parallel, 50 nodes
  • 40,000 nodes = 800 million edges (pairwise)
  • 40,000 ! Potential relationships
  • Big data or big relationship domain

• Network analysis: Best in parallel
  • Only 3% of entire genome forms complexes

• Holland Computing Center: Firefly 1150 8-core cluster – from weeks to hours/minutes
The Need for HPC

Correlation Network Creation

Time in hours

2 weeks

# Processors
Network Filters

Design a network filter and obtain a sub-network of the original network such that:

• It maintains the important stuff – signal
• Remove unimportant stuff – noise
• Maintain network elements of biological relevance
• Uncover new ones
Network Filters

• Chordal graph sampling
  • Keep triangles in expression graphs
  • Remove large cycles, extra edges
  • Keep clusters, identify new clusters

• Spanning tree sampling
  • Keep high degree nodes (maybe?)
  • Remove up to 50% of edges
  • Enhance identification of lethal nodes

• Planar Graph Sampling
  • Maintains key planar structures – like pathways
  • Does not maintain clusters
  • Remove up to 60% of edges
Chordal Graph Sampling

Goal: Develop a parallel network sampling technique that filters noise, while preserving the important characteristics of the network.

✓ Maximal Chordal Subgraph
  ➢ Spanning subgraph of the network
  ➢ No cycles of length larger than three

✓ Properties of Chordal Graph
  ➢ Preserves most cliques and highly connected regions of the network
  ➢ Most NP hard problems can be solved in polynomial time
  ➢ Complexity of finding maximal chordal subgraphs: $O(|E|*\text{max}_\text{deg})$
Need to Maintain Key Structures
Proposed Approach

✓ Create networks from publicly available data
  ➢ Aging mice gene expression data –
  ➢ Young vs. middle-aged mice

✓ Test method on networks

✓ Assess results by examining biological relevance of network structures
  ➢ Clusters enriched with function (Gene Ontology)
  ➢ Do we maintain clusters and function in sampled graphs?
  ➢ Do we find new functions in sampled graphs?
Hypothesis

- Hypothesis $H_0$: Given a graph $G$ representing a correlation network, maximal chordal subgraph $G_1$ will maintain most of the highly dense subgraphs of $G$ while excluding edges representing noise-related relationships in the network.

  - $H_{0a}$ - Key functional properties found in the clusters of unfiltered networks $G$ are maintained in the sampled networks $G_1$

  - $H_{0b}$ - New clusters with biological function are uncovered. Functional attributes previously lost in noise can now be identified.
Gene expression network of the hypothalamus of a mouse brain with components highlighted (A-D) and respective chordal graphs shown (A’-D’).
Identification of New Clusters

<table>
<thead>
<tr>
<th>Network</th>
<th>Conserved clusters</th>
<th>Newly identified clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young Mouse RCM Ordering</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Young Mouse BFS Ordering</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Middle Aged Mouse BFS Ordering</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Middle Aged Mouse RCM Ordering</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Tutorial Outlines

- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics and Heath Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects of Network Analysis
- Next Steps – where to go from here?
Challenges

• **Technical Challenges:**
  – The need to analyze existing data not just collect data Are they all accurate? Complete?
  – Aggregation versus no aggregation
  – Ideal level of analysis - granularity
  – How results can be verified? Validated?

• **Philosophical Challenges:**
  – Collaboration
  – Work on the correct problem
  – Need for genuine translational work
SWOT Analysis

• Good News:
  – We have the data
  – We have a number of useful tools
  – We have the talent

• Challenges:
  – Can we trust the results: If we torture data sets long enough, they will confess
  – New models for data integration
  – Collaboration and Interdisciplinary work
Next Generation Research

- Next Generation Tools or Devices need to be Intelligent, Collaborative, and Dynamic
- Biomedical scientists, Bioinformatics researchers and Healthcare providers need to work together to best utilize the combination of tools development and domain expertise
- The outcome of collaboration has the potential of achieving explosive results with significant impact on human health and overall understanding of biological mysteries
- Advanced computational tools, cloud and HPC technologies are critical to the success of the next phase of Biomedical research but again the integration needs to happen at a deeper level