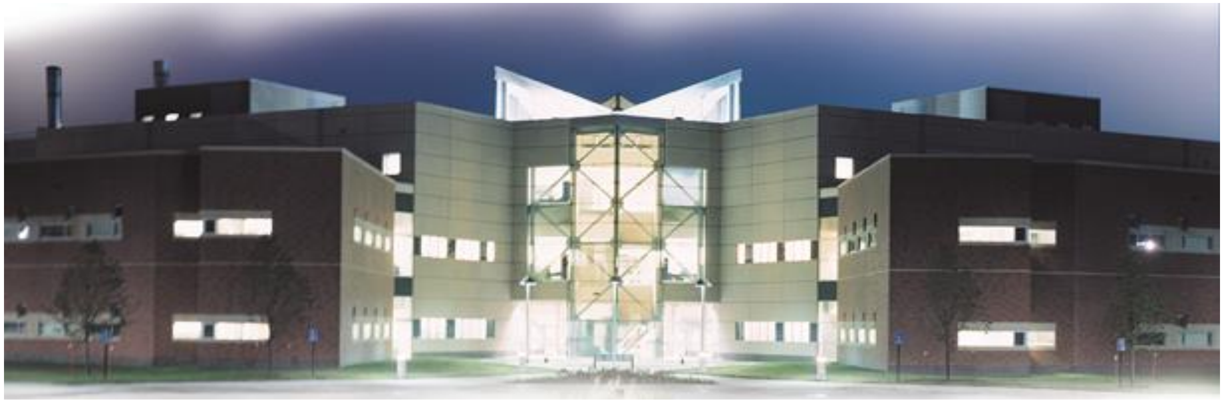
University of Nebraska at Omaha

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

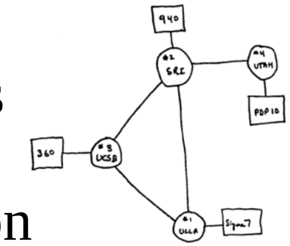
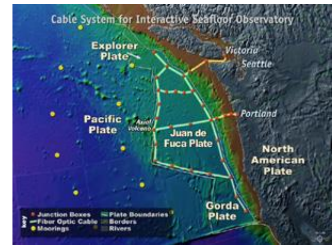
- ***Scientific Data-Driven Revolution: An Overview***
- Big Data Analytics and Heath Monitoring
- Wireless Sensors and Mobility Analysis for Healthcare
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- Civil Infrastructure and Data Analytics
- Technical Implementation Aspects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

Scientific Research will never be the same

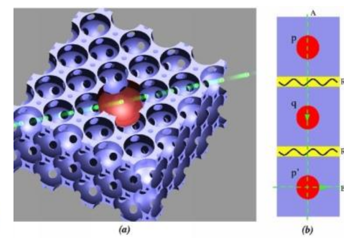
- IT is changing many scientific disciplines
- So much relevant data is currently available
- The availability of data shifted many branches in sciences from pure experimental disciplines to knowledge based disciplines
- Incorporating Computational Sciences and other branches of sciences is not easy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

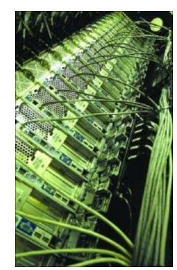
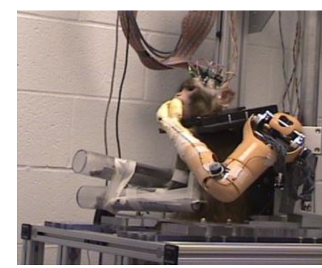
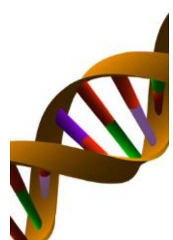
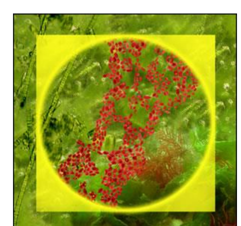
The Future is Full of Opportunity

- 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 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?

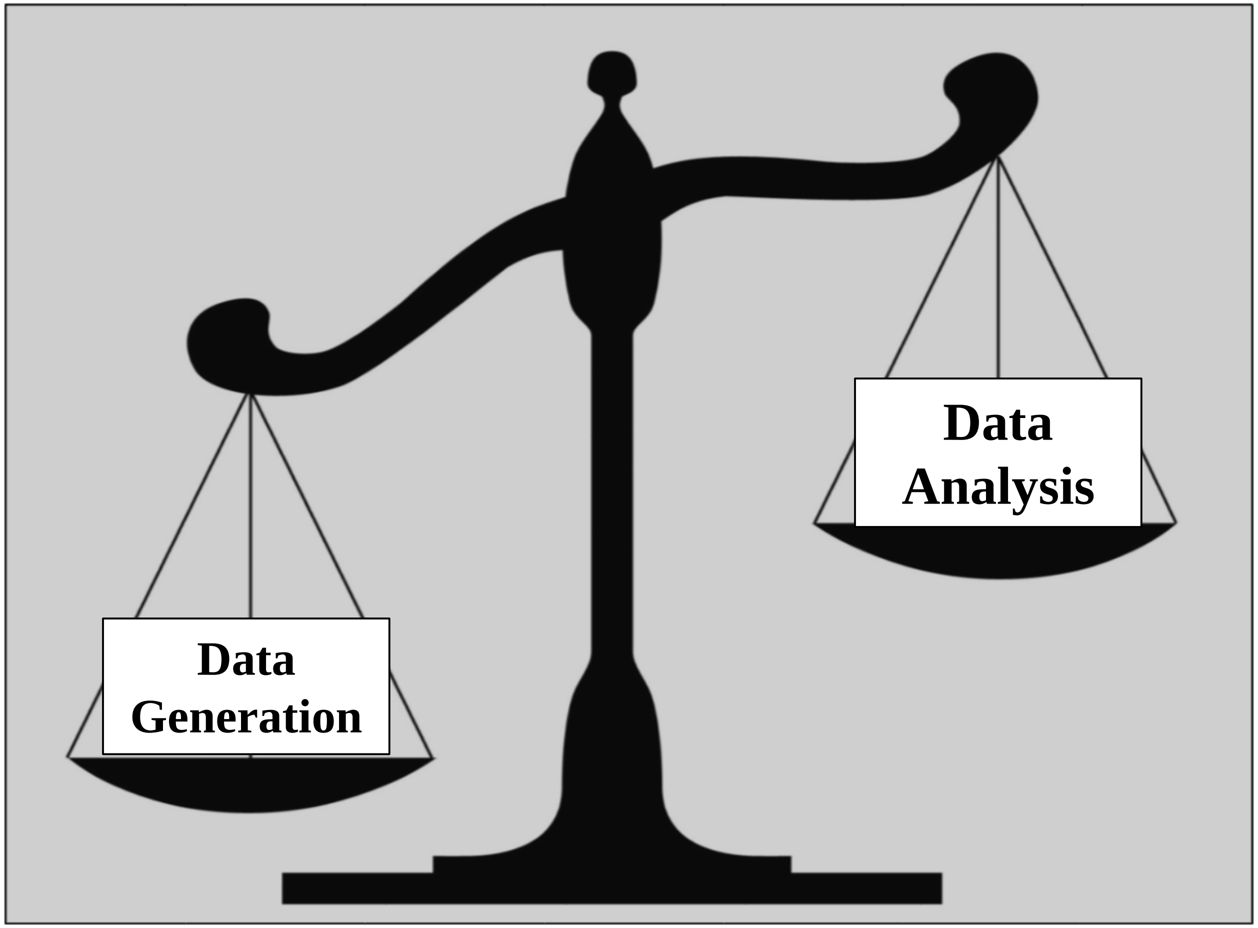
Challenges

- 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?

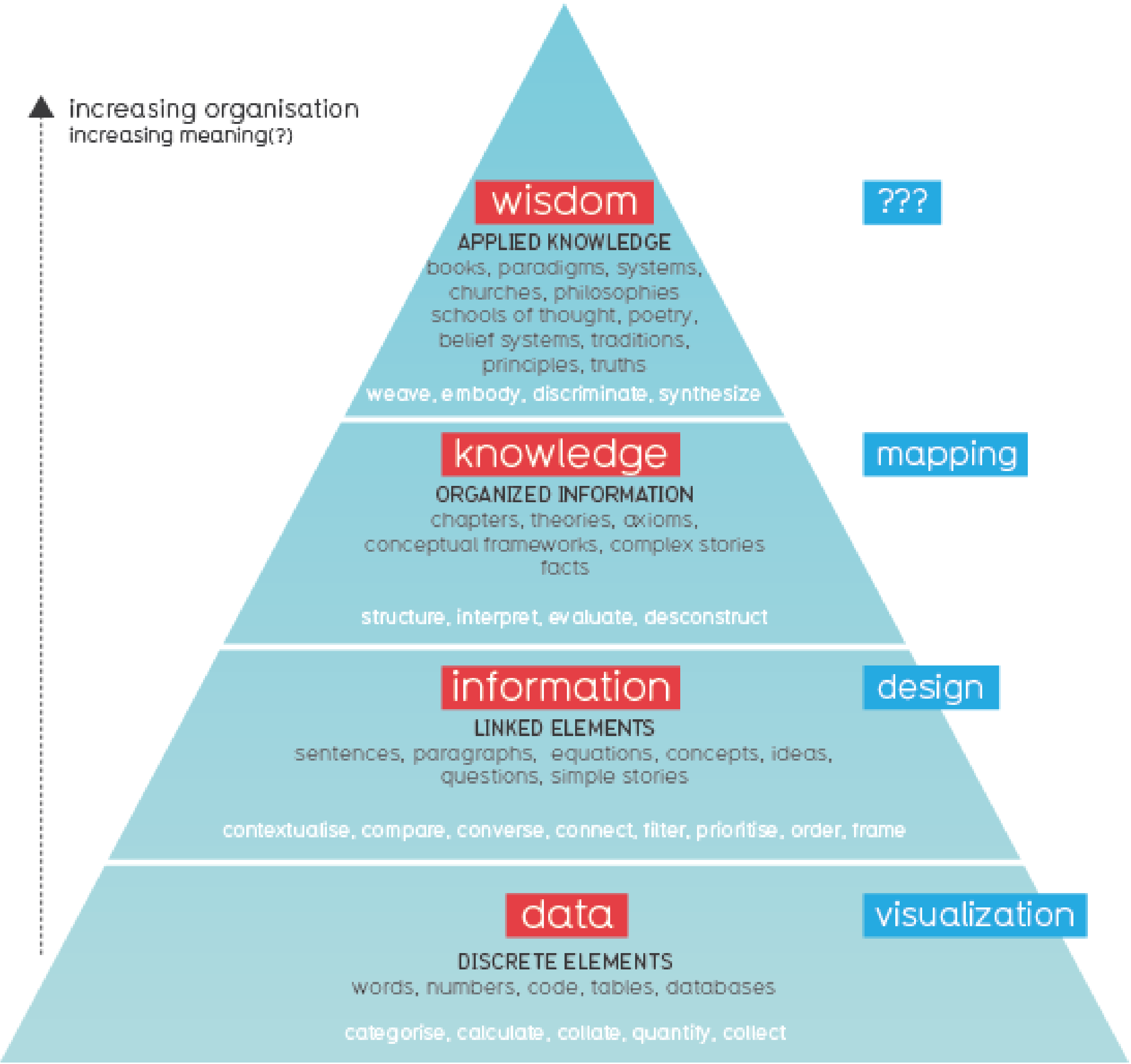
Data Generation vs. Data Analysis/Integration

- 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



http://www.ritholtz.com/blog/wp-content/uploads/2010/11/data_info_knowledge_wisdom.png

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 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

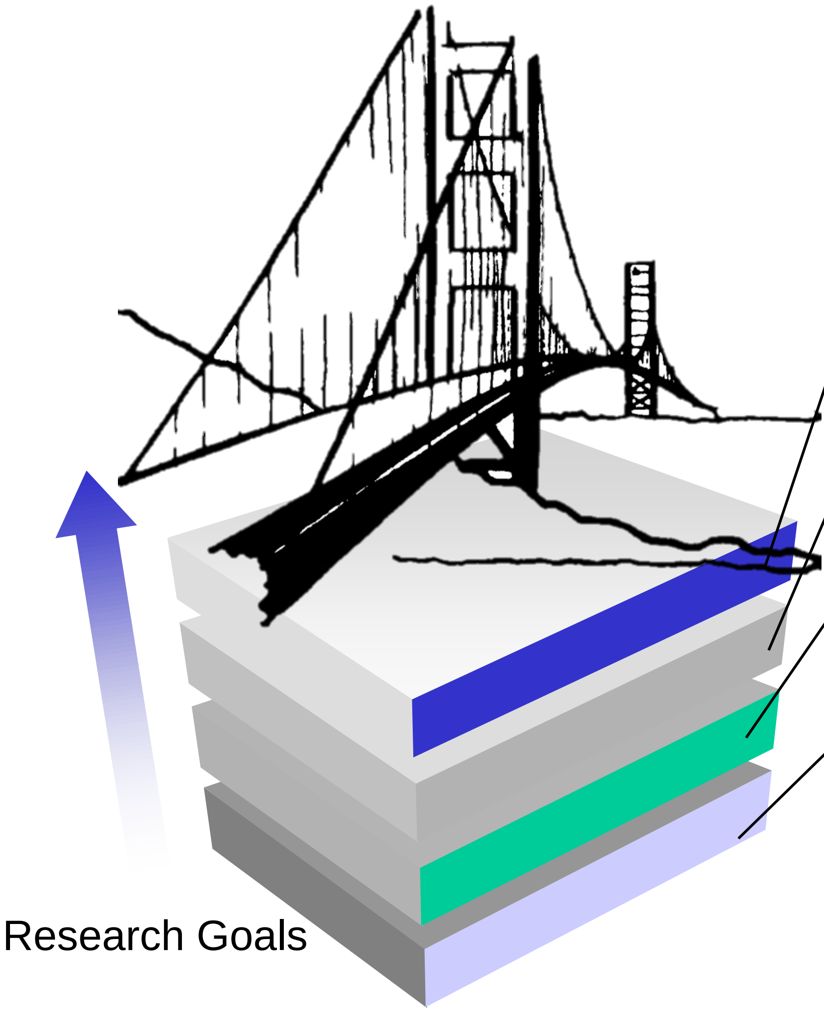
- Numerous incidents of structural failures

- 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

UNO BIOINFORMATICS

- 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?

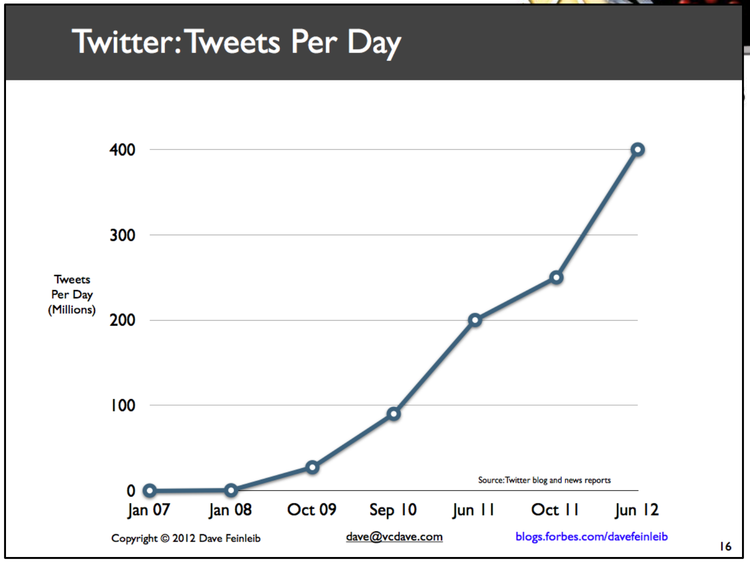
What *isn't Necessarily* Big Data?

- 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

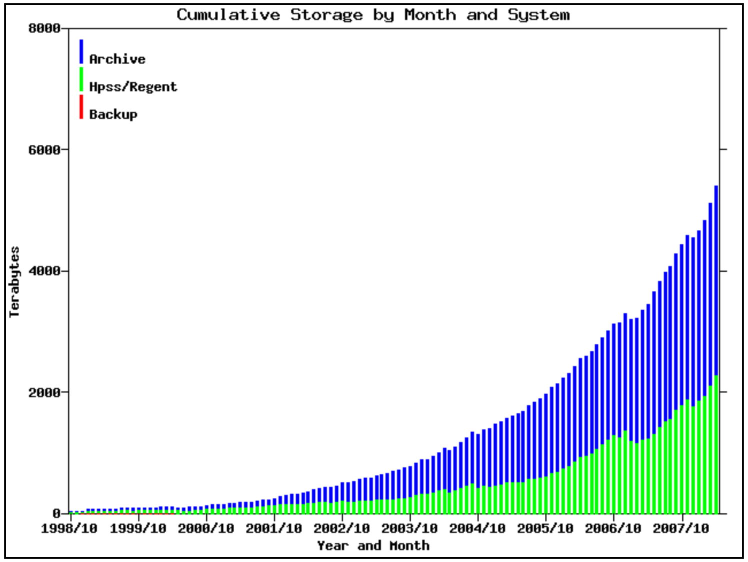
“Big Data”

- ***Big data:*** “Any data too big to be handled by one computer” – Scientist John Rauser[1]
- 90% of worlds 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]

Forbes: 400 million Tweets/Day

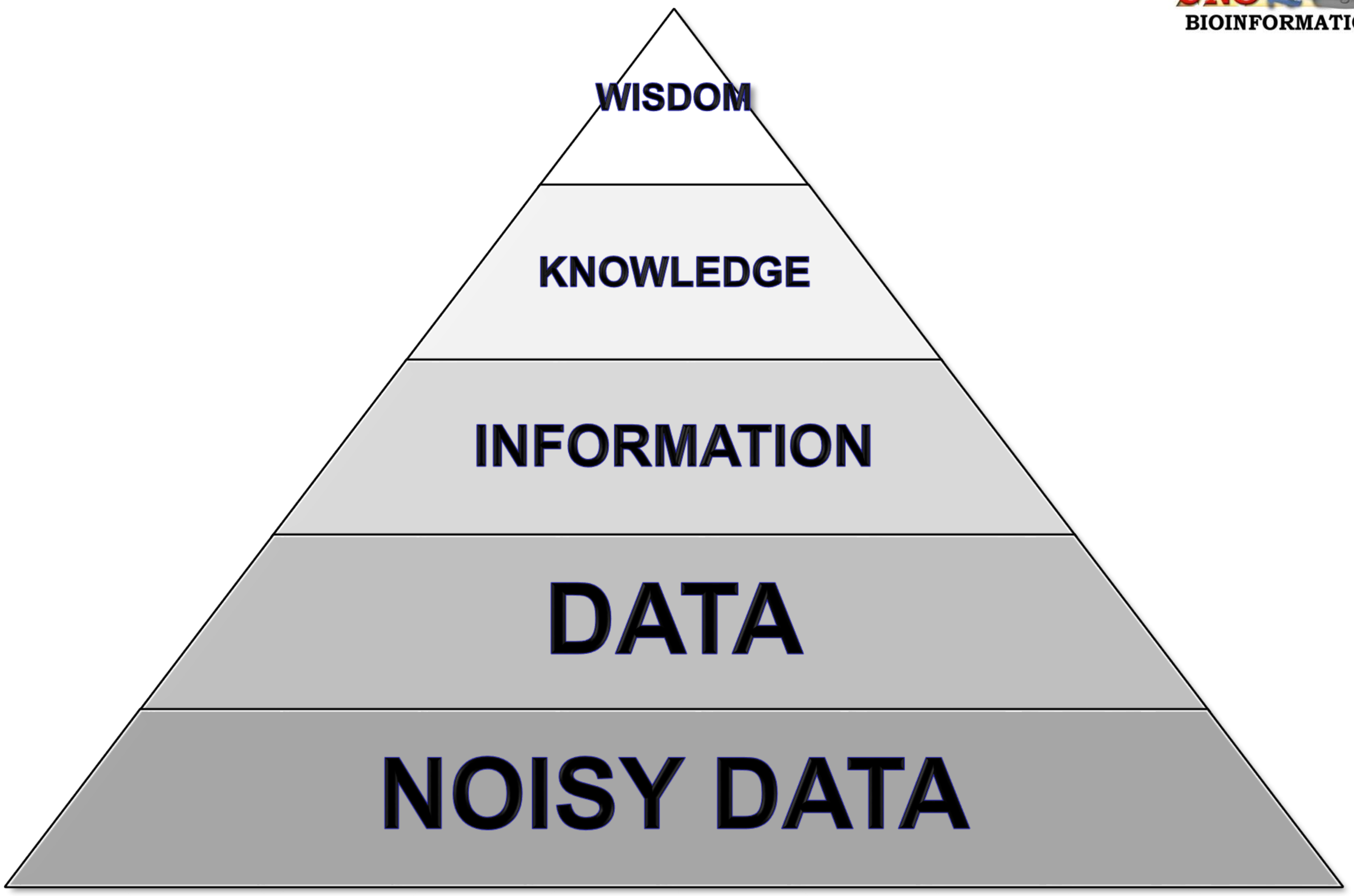


NERSC: 6PB of data since 1998



[1] http://www.networkworld.com/news/2012/051012-big-data-259147.html
[2] http://www-01.ibm.com/software/data/bigdata/
[3] http://www.economist.com/node/15557443?story_id=15557443

Updated Model



The “Systems” Approach

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

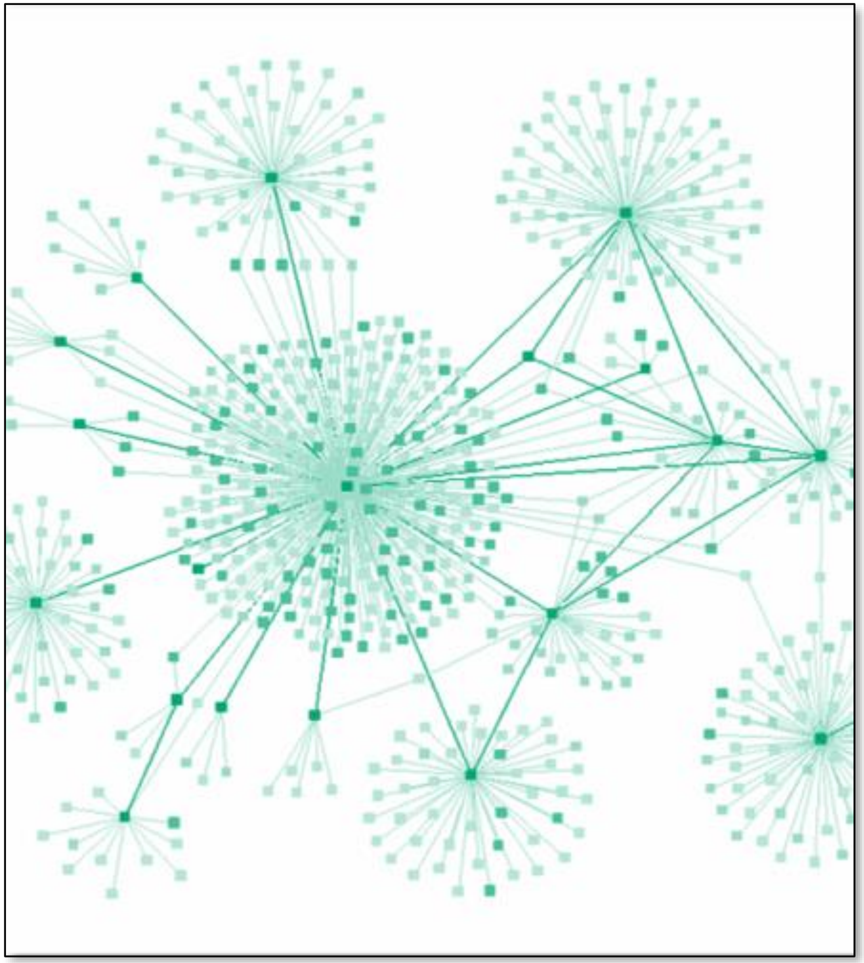
- Discovery
  - Examine changes in systems
    - Normal vs. diseased
    - Young vs. old
    - Fast vs slow
    - Stage I v. State II v. Stage III v. Stave IV

The Network "Graph" Model

- A network represents elements and their interactions

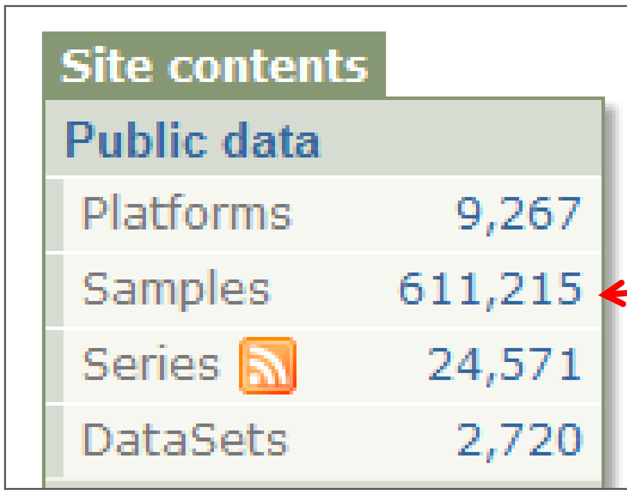
- Nodes → elements
- Edges → interactions

- Can represent multiple types of elements and interactions



Why Networks?

- Explosion of biological data



Site contents

Public data	
Platforms	9,267
Samples	611,215
Series	24,571
DataSets	2,720

Each sample can have over 40,000 genes

- 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

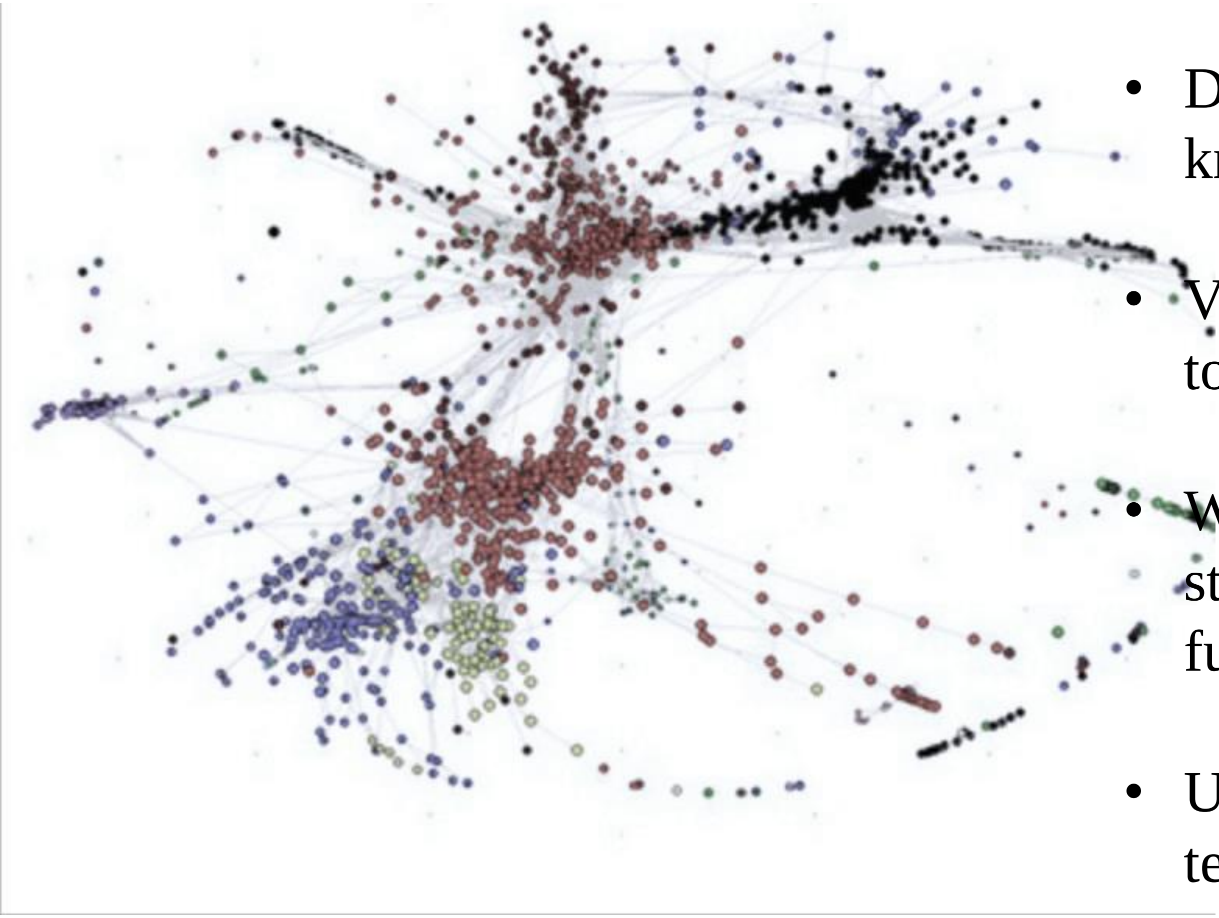
Structure and function have been found to be related in social, biological, co-citation, transportation (etc.) networks[1,2].

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

1. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509-512. doi: 7898 [pii].
2. Barabasi AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. *Nat Rev Genet*. 2004;5(2):101-113. doi: 10.1038/nrg1272.

Synthetic Lethality

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

UNO BIOINFORMATICS

- 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

Individual's health is a big data problem

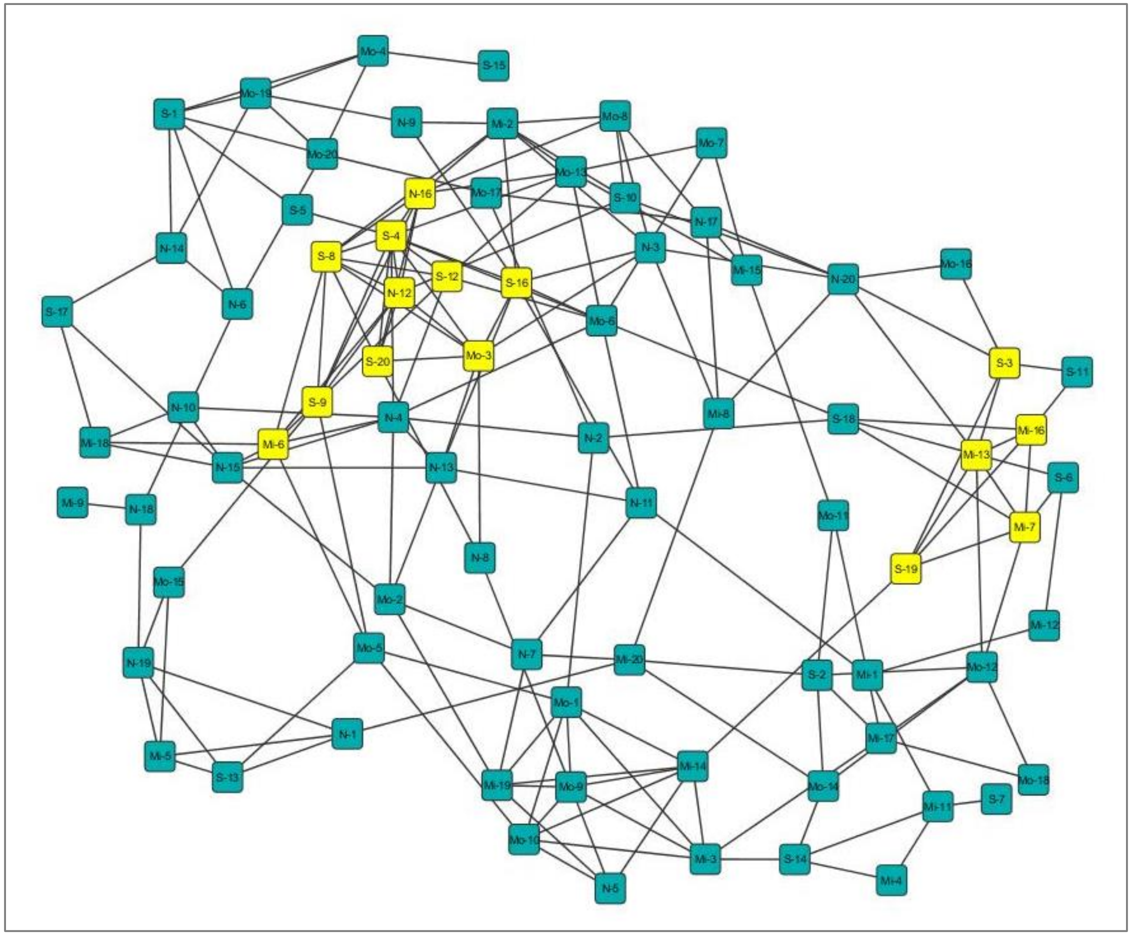
"The future of biomedical research is about assimilating data across biological scales from molecules to populations. **As such, the health of each one of us is a big data problem**. Ensuring that we are getting the most out of the research data that we fund is a high priority for NIH."

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 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?

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

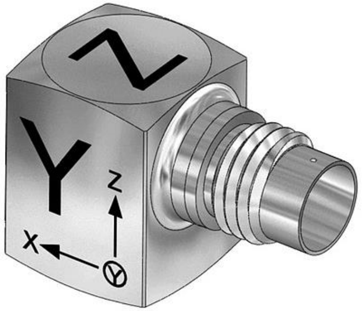
- Laboratory setting
- Real-world setting
- Self-reported data collection method
- Using monitoring devices, sensors and accelerometers
- GPS
- Gyroscope

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

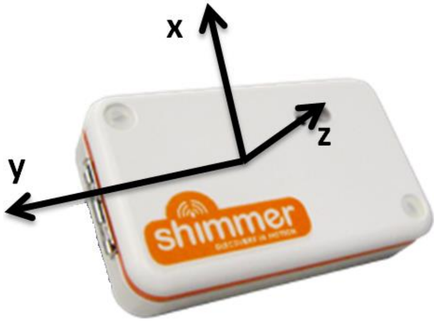
Pedometers versus Accelerometers



- 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

UNO BIOINFORMATICS

Table 2 Selected moderate and vigorous activities, physiologic pathways, and health outcomes[a]

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

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

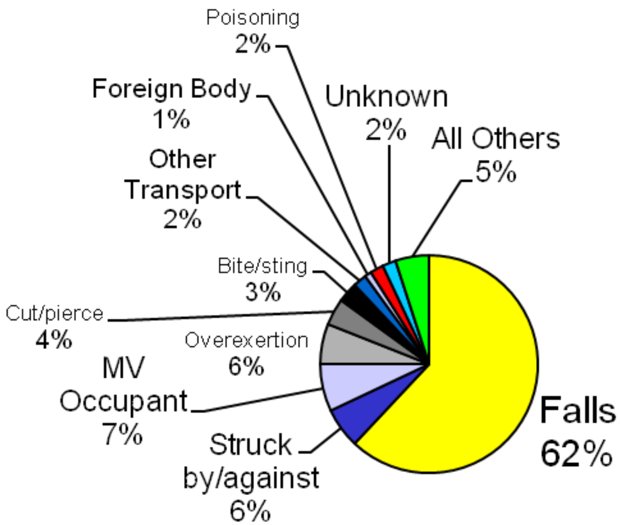
Wireless Networks in Aging

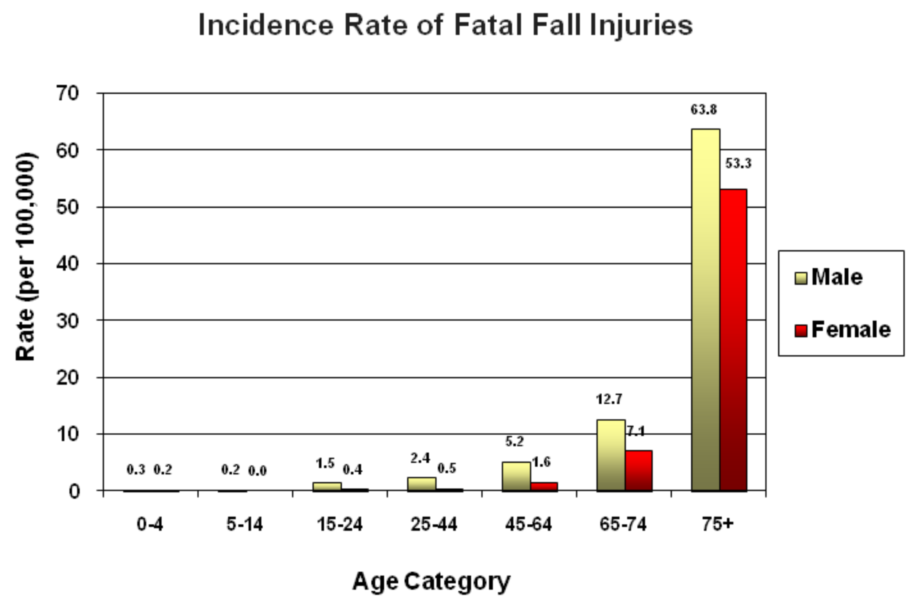
- Correlation between mobility and health level
- Monitoring mobility levels
- Aging of cells and aging of systems
- Collaboration between Bioinformatics group, Wireless Networks group and Decision Support Systems group

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



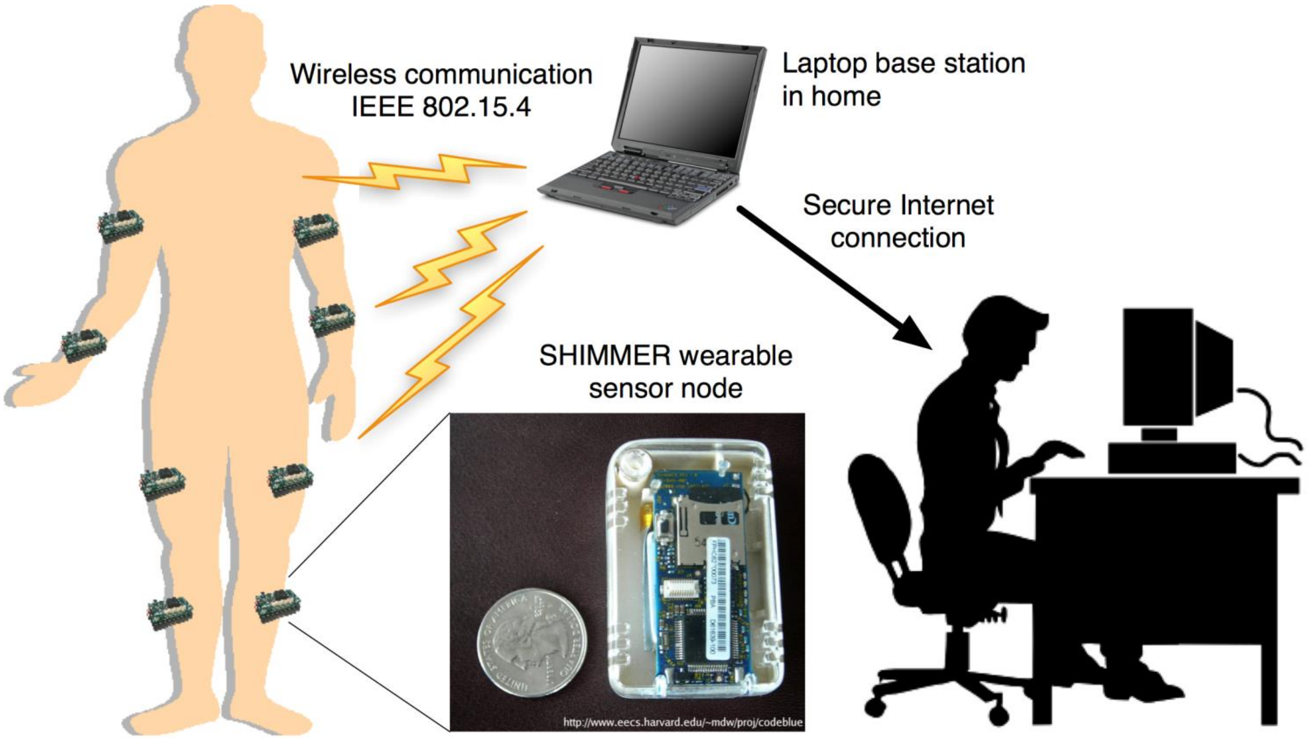


- Approximately 78% → 65 years and older.
- falls – leading cause of
  - injury deaths
  - injuries and trauma
- The risk of falling increases with age.

Slide: Courtesy of Dr. Nick Stergiou

Wireless Sensor Based Mobility Monitoring

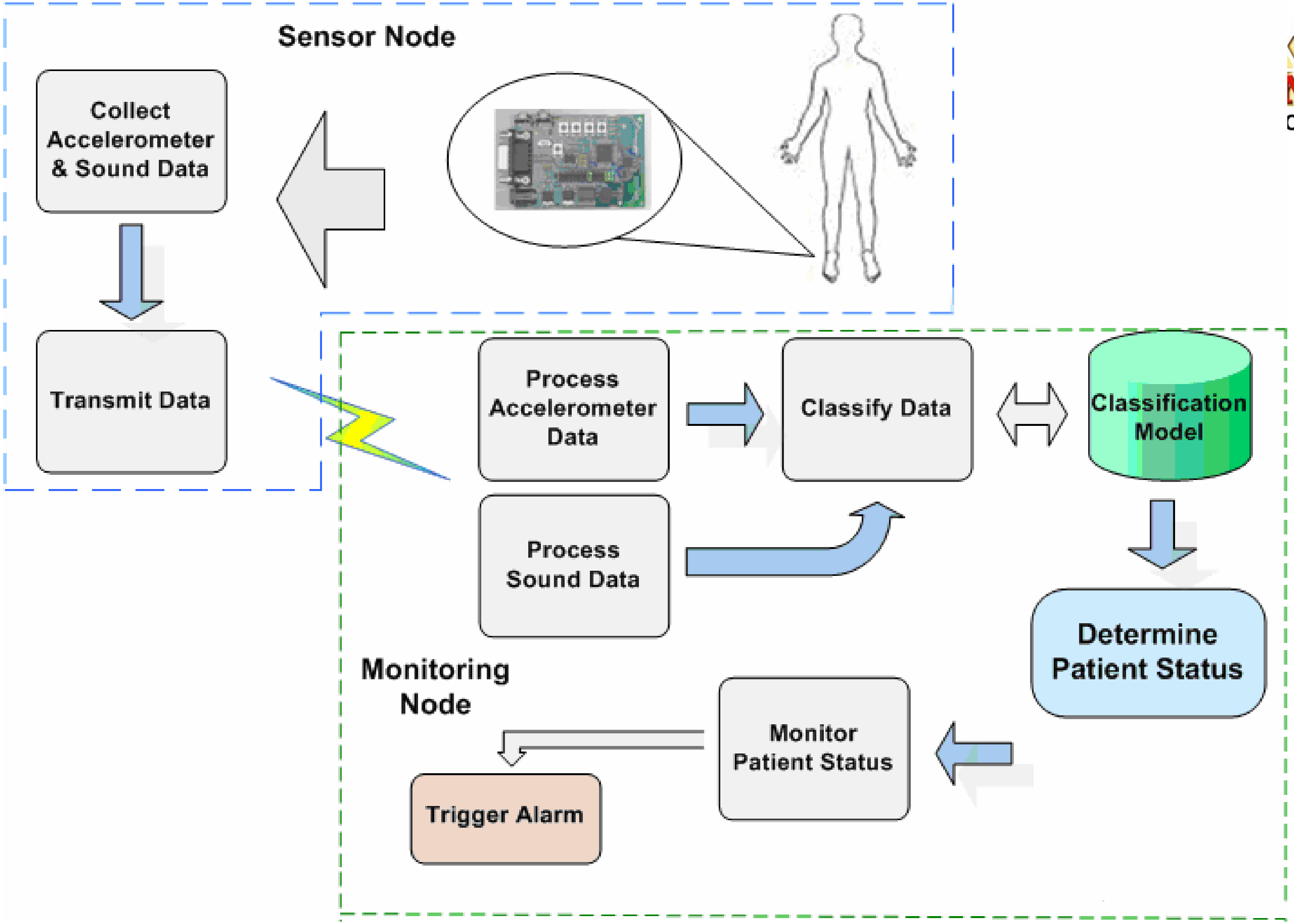
- Inexpensive
- Comfortable
- High mobility
- Simple



http://fiji.eecs.harvard.edu/Mercury

Four Phases of the Project

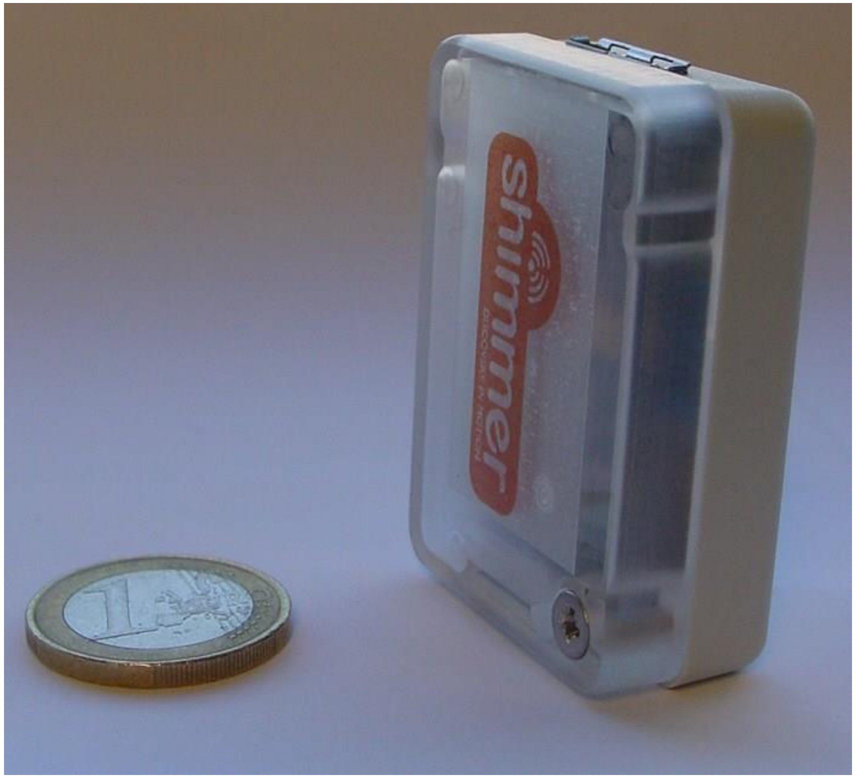
- Phase I: Fall Detection
  - achieved over 95% of fall detection rate
- Phase II: Classification of ADLs (Activities of Daily Living)
  - Running, Walking, Stair Climbing, Jumping, Stair Climbing, Standing, Sitting, and Lying, …
- Phase III: Construction of Mobility Profiles
- Phase IV: Fall (major 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)



(shimmer-research.com)

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

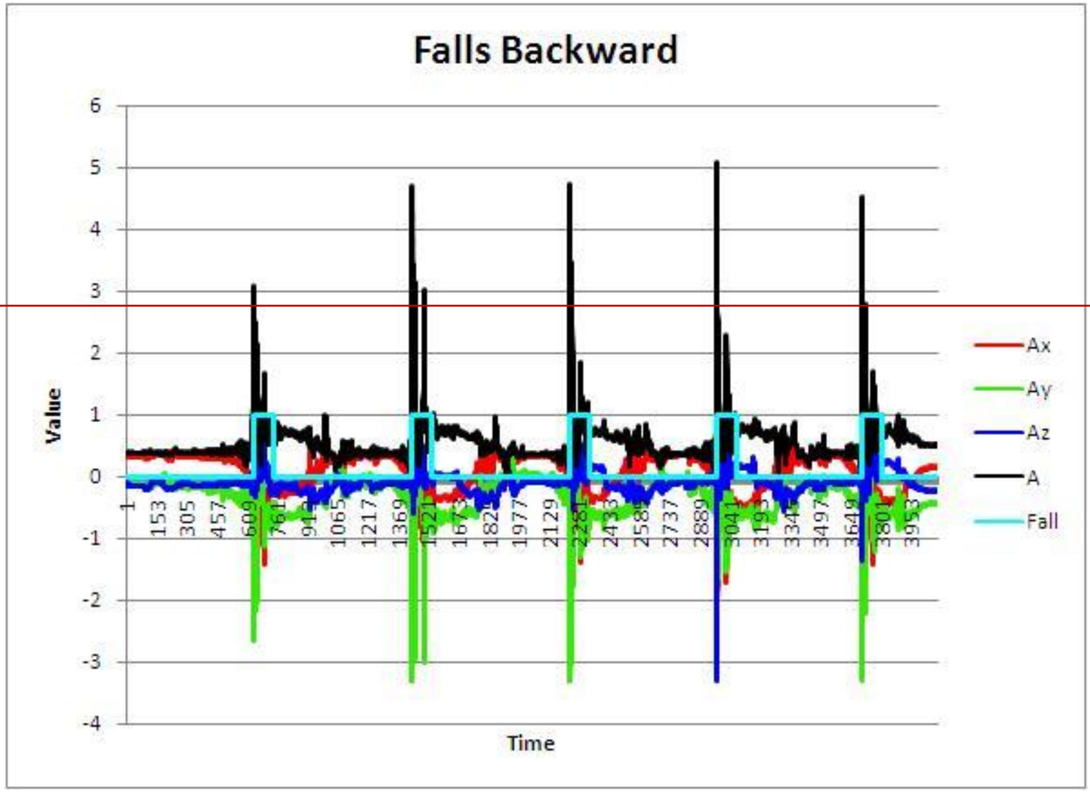
Phase I: Fall Detection

Accelerometer-based fall detection

UNO BIOINFORMATICS

- Determine an acceleration threshold.
- Detect fall.

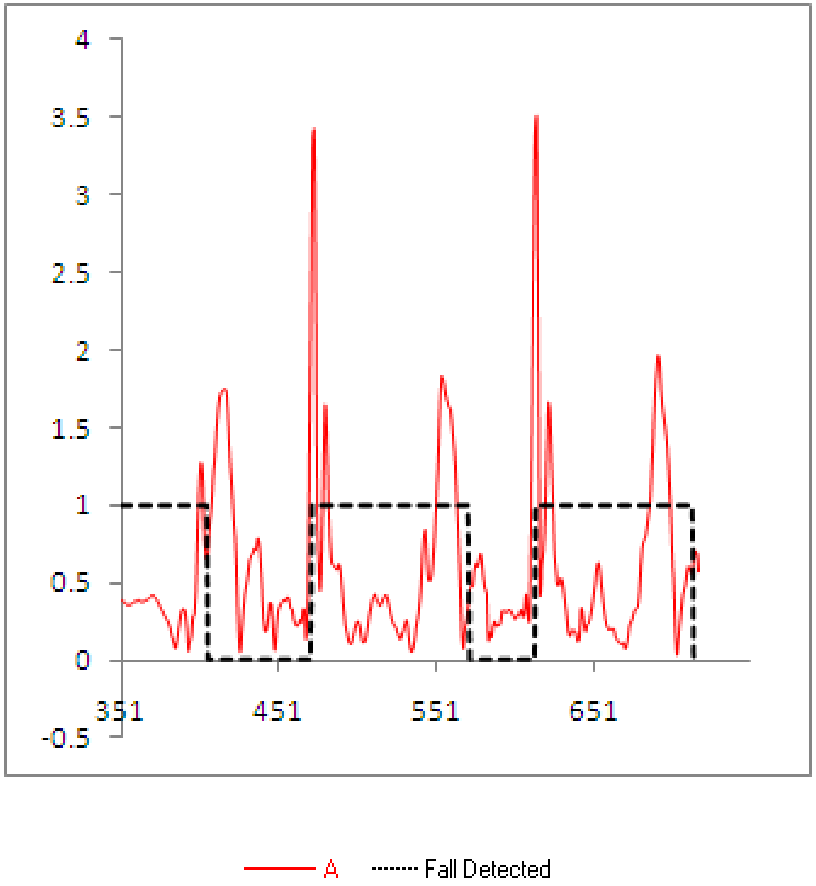
Threshold



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.

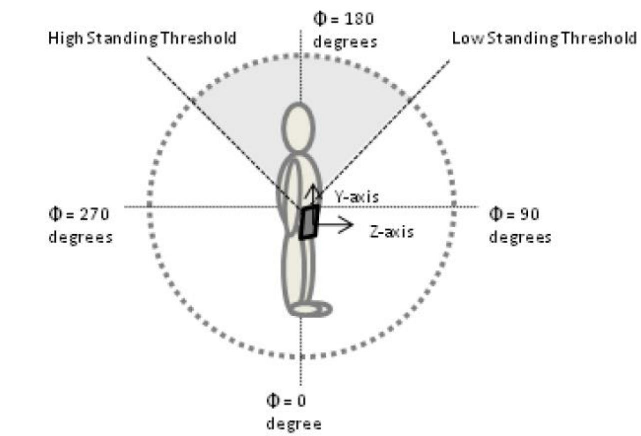
- Key Metrics
  - Inclination Angle
  - Standard deviation
  - Skewness
  - Signal Magnitude Area

Phase II: Classification of ADLs

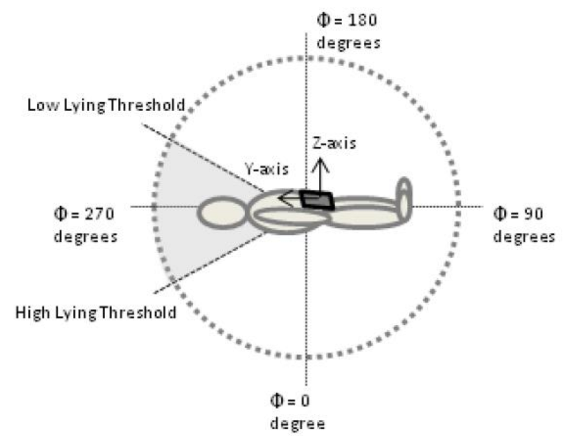
UNO BIOINFORMATICS

- 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



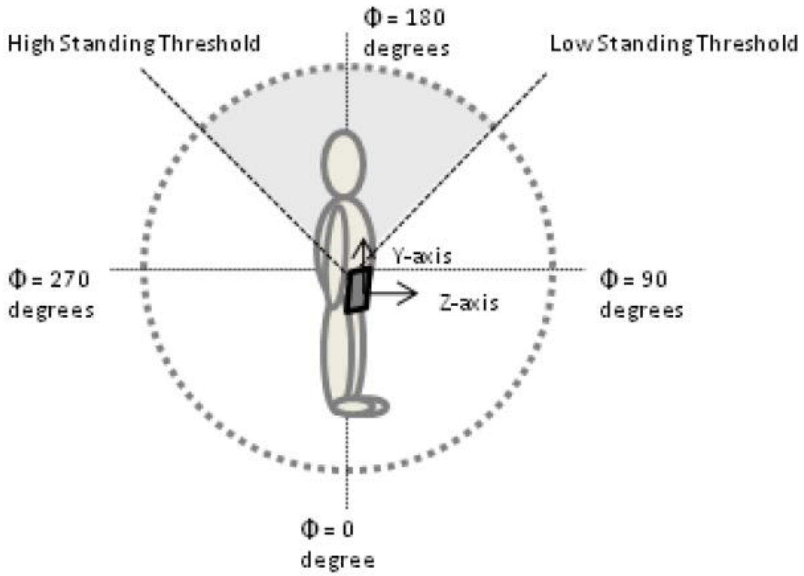
Standing Position



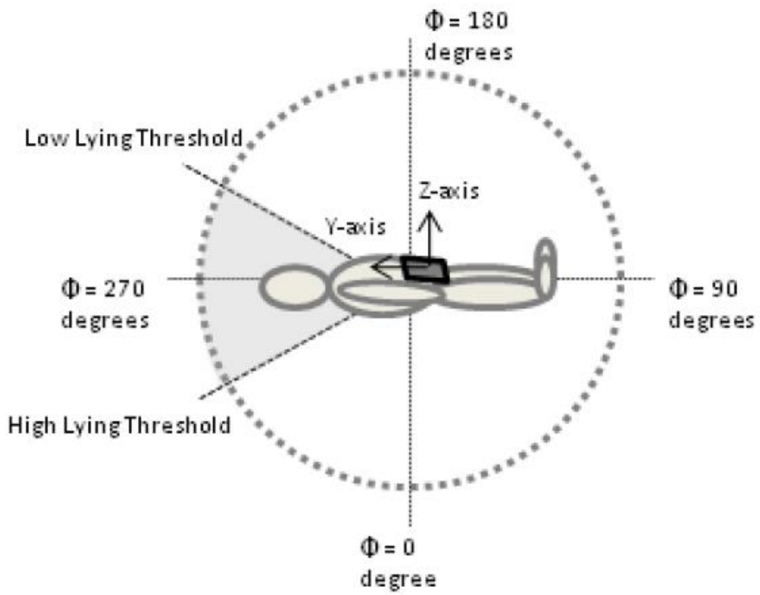
Lying Position

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.



Standing Position



Lying Position

- Can now differentiate standing, sitting, and lying.

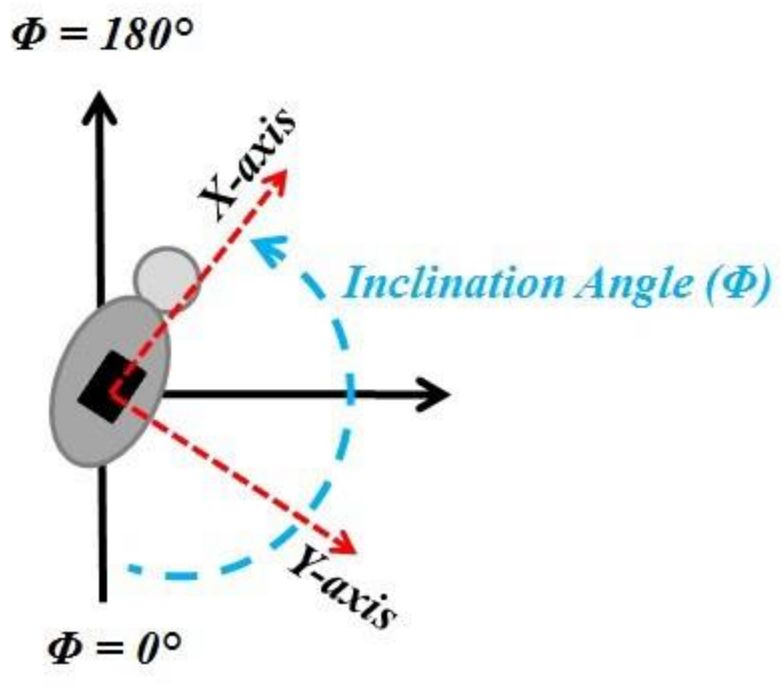
Activity Classification

- 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

$$\Phi = \arctan \frac{A_y}{A_x}$$

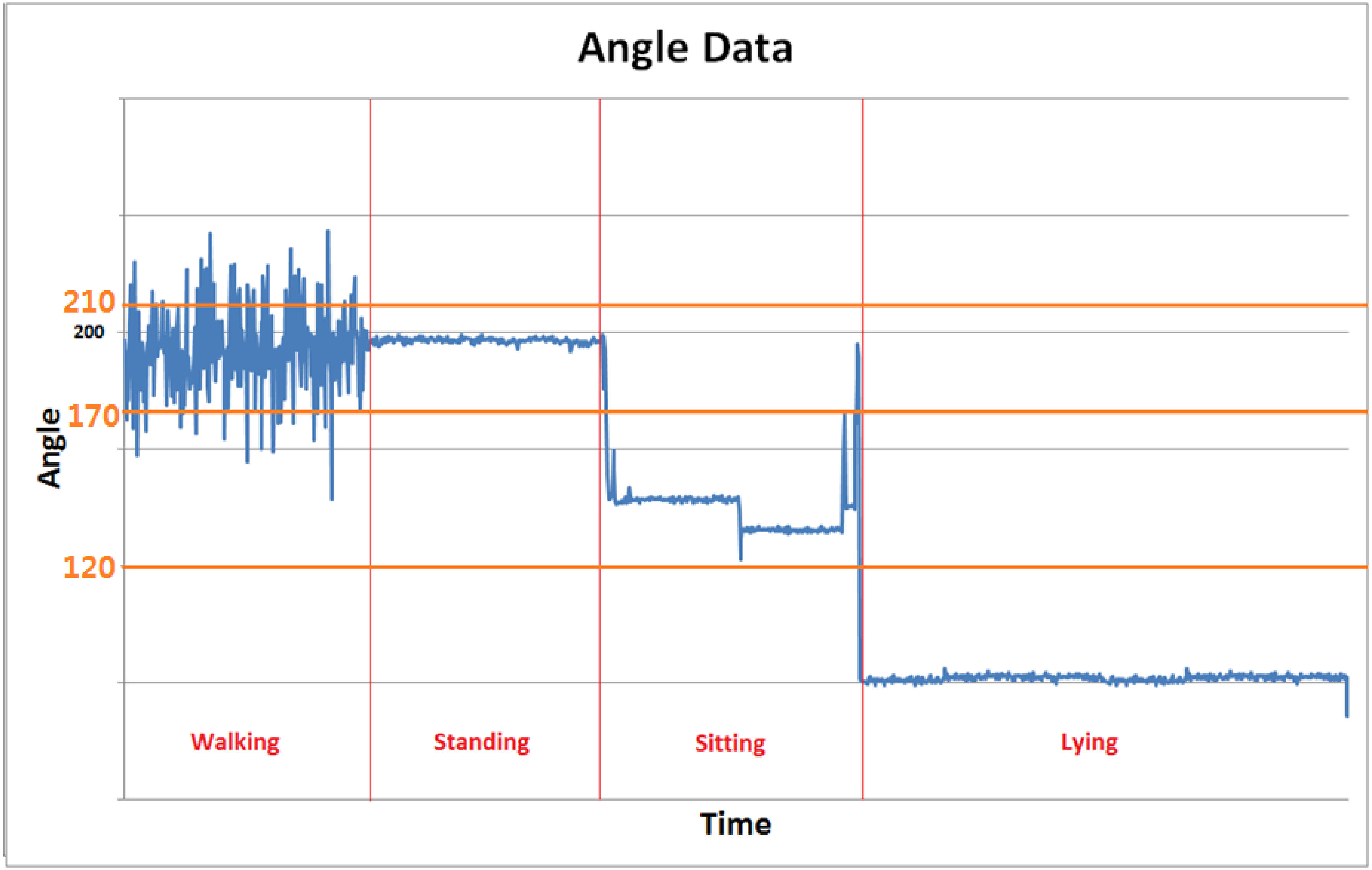
A_x: acceleration value of x-axis

A_y: acceleration value of y-axis



Inclination Angle

UNO BIOINFORMATICS



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.

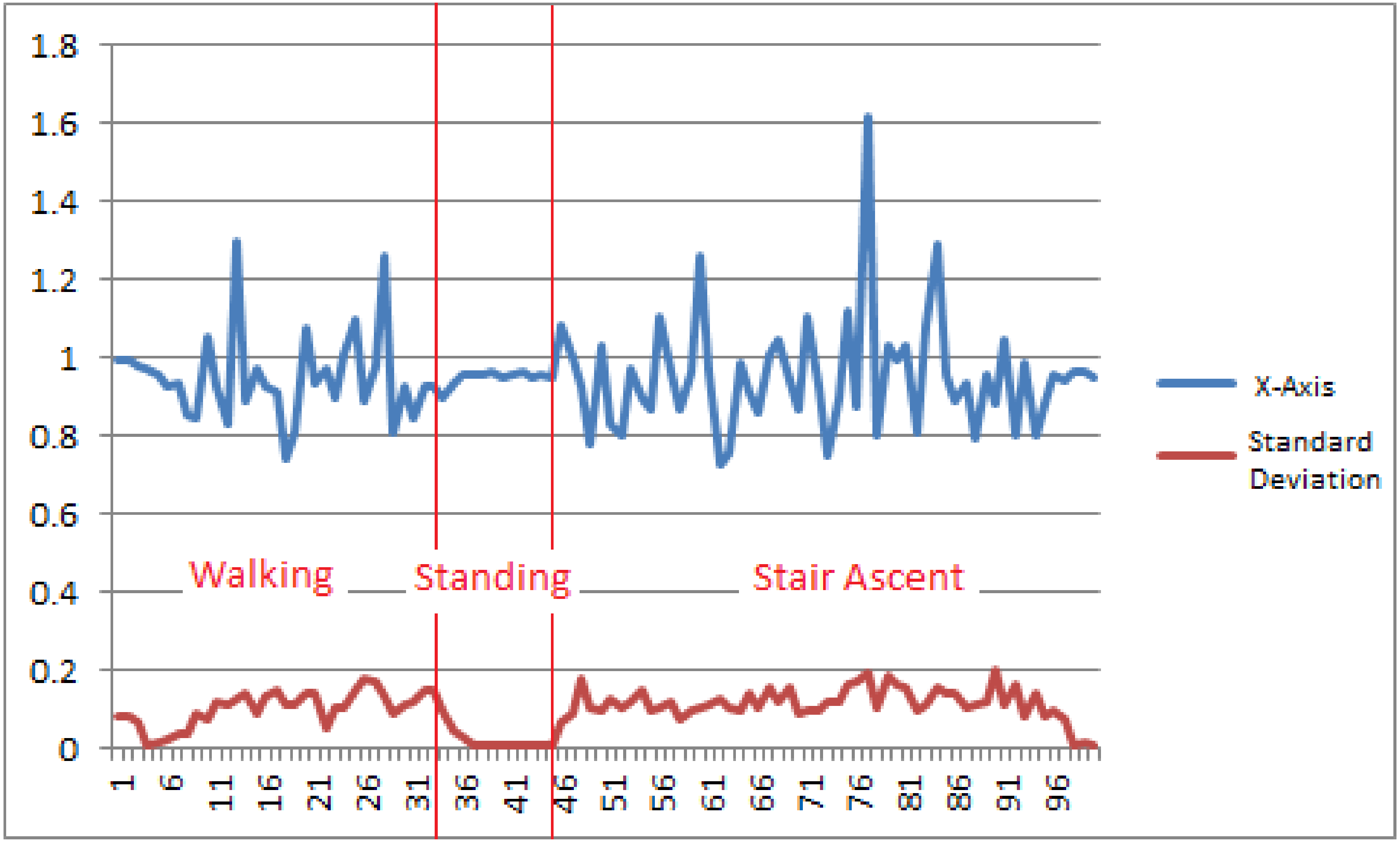
- When used with angle measurement, can now differentiate standing from walking/running.

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

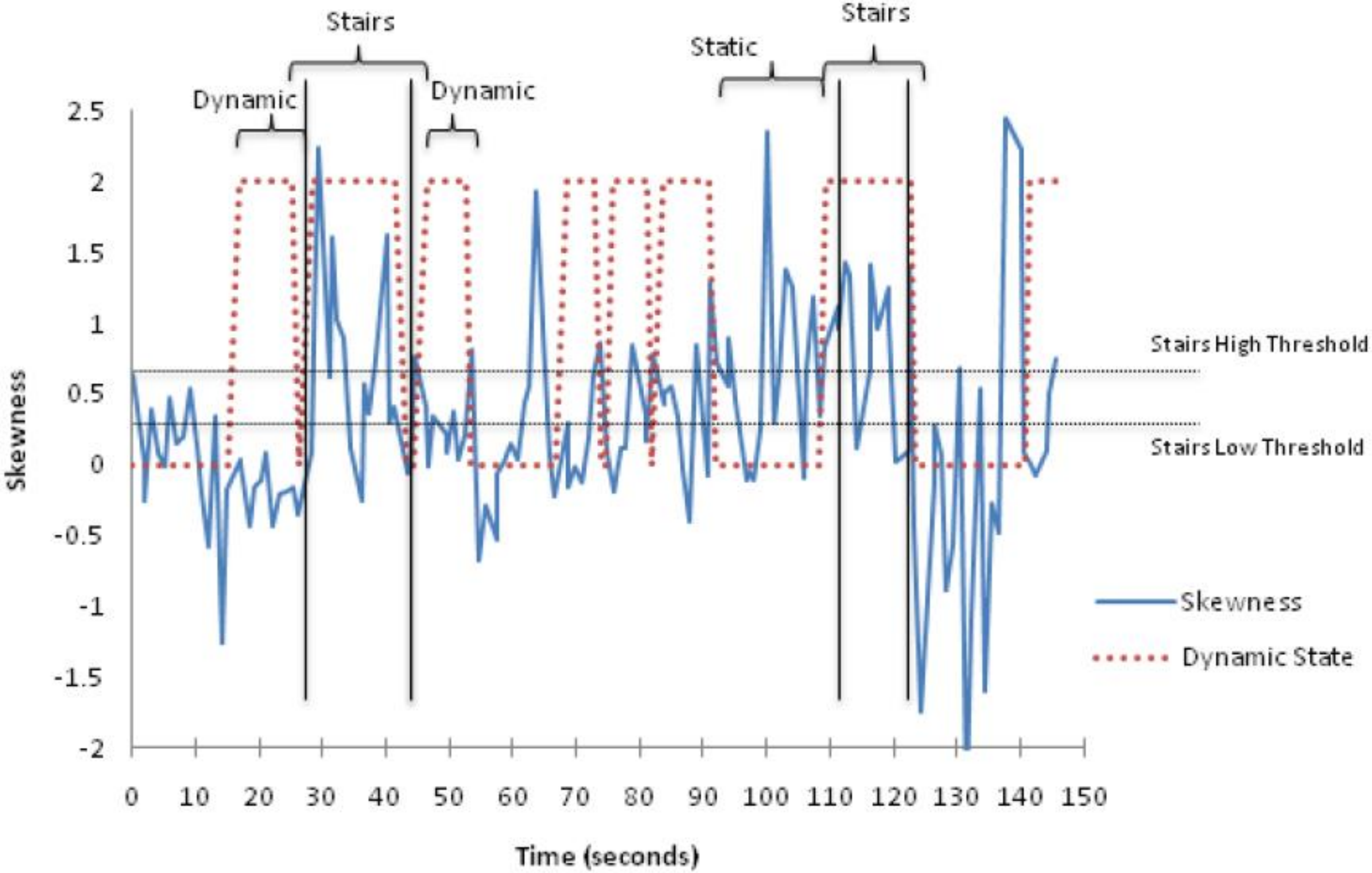
- Skewness helps in determining if the current mobility state is going up or down stairs.
  – Skewness measures the asymmetry of the distribution of x-axis acceleration values. It is assumed that going up/down stairs will produce data that has greater asymmetry than walking.
- When used with angle and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



G. Hache, E. Lemaire, N.Baddour, “Development of wearable mobility Monitoring System, “ in Proc. Can. Med. Biological Eng. Conf., Calgary, Canada, May 2009.

Signal Magnitude Area

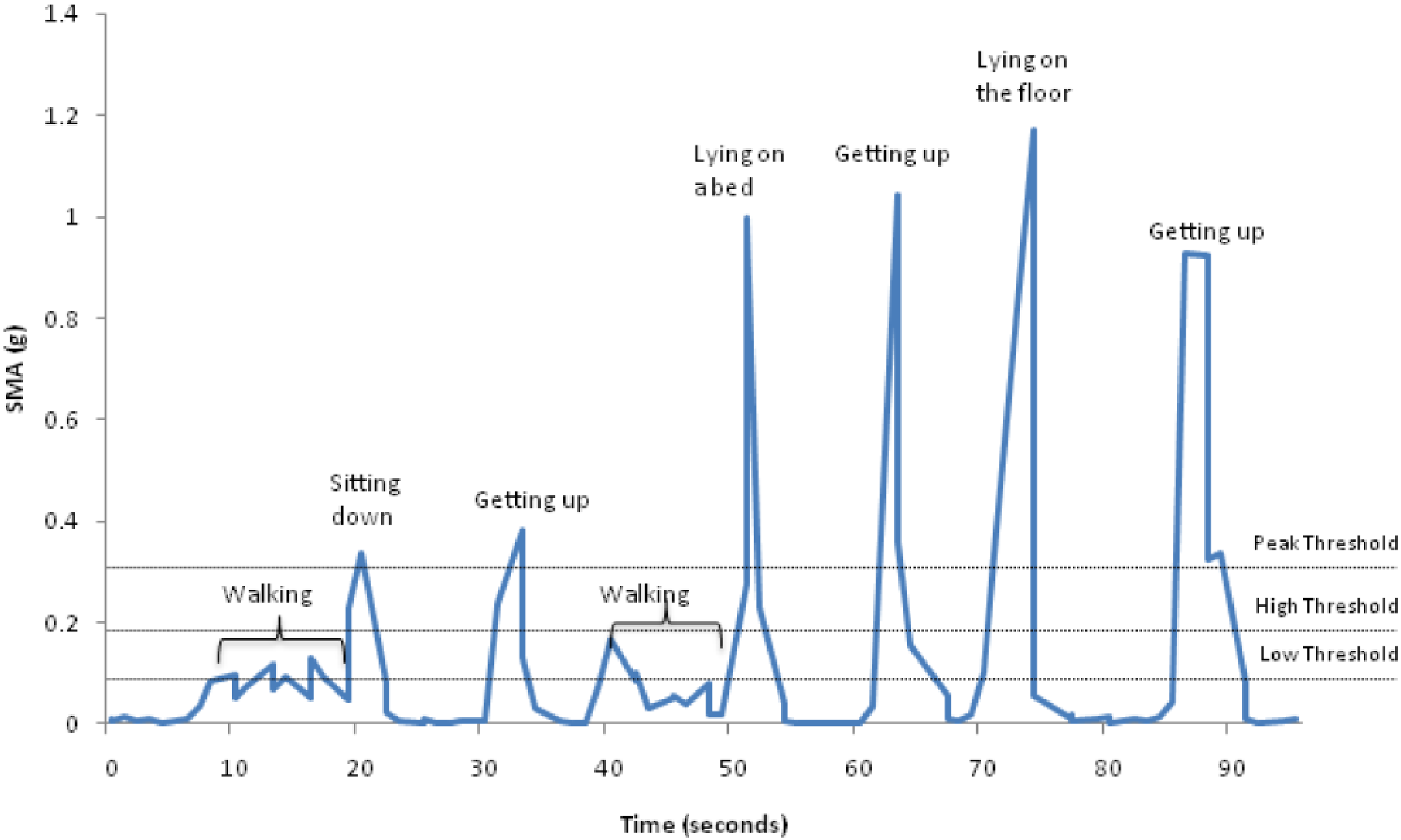
- Measures amplitude and duration variation in the acceleration signal.
  - Assumed that amplitude and duration variation will be greater when the intensity of the activity changes. As such, SMA values will be greater when changing state as opposed to not changing state. Ex: getting out of bed vs. walking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

$$SMA = \frac{1}{t}\left(\int_{t=0}^{T} |a_x| dt + \int_{t=0}^{T} |a_y| dt + \int_{t=0}^{T} |a_z| dt\right)$$

Signal Magnitude Area (SMA) of acceleration signals versus Time



G. Hache, E. Lemaire, N.Baddour, "Development of wearable mobility Monitoring System, " in Proc. Can. Med. Biological Eng. Conf., Calgary, Canada, May 2009.

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-there-of 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

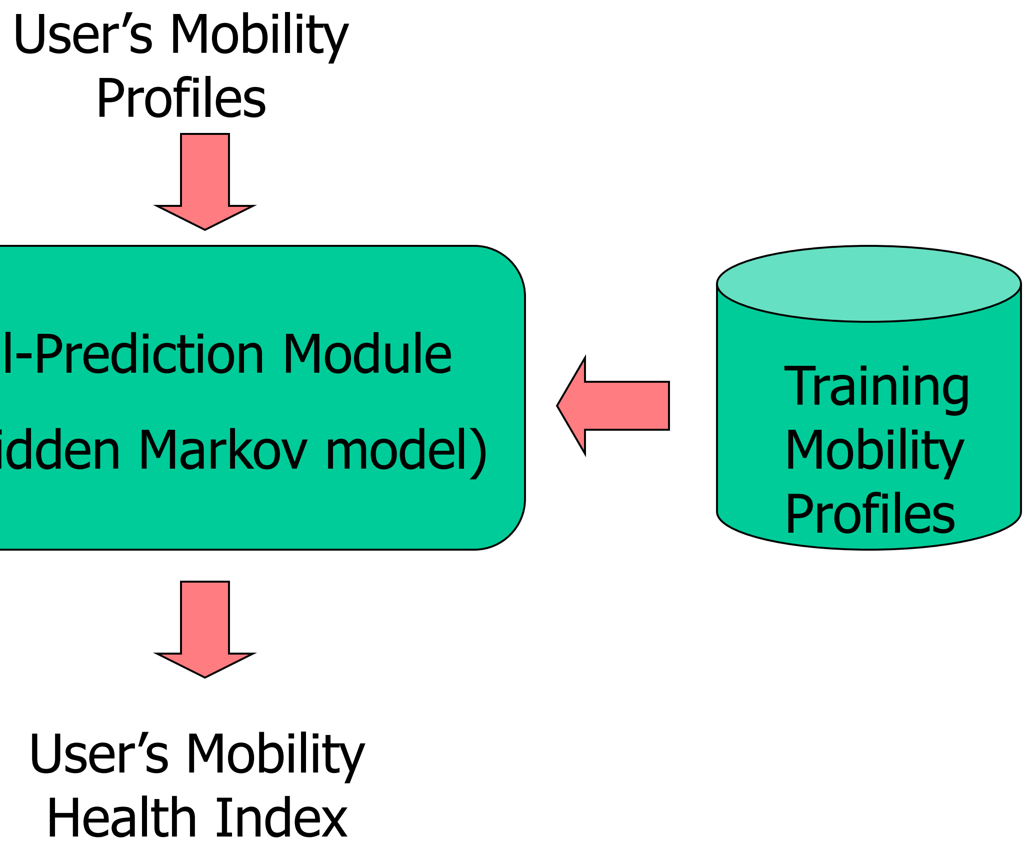
Activity	Place	Time (second)	
		Subject I	Subject II
Walking	Outdoor	360	300
Standing	Outdoor	20	20
Running	Outdoor	30	15
Standing	Outdoor	30	120
Walking	Outdoor	300	300
Sitting	Indoor	360	340
Walking	Indoor	30	45
Lying	Indoor	72	60
Total time		1200	1200

Profiles: Experiment Results

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.

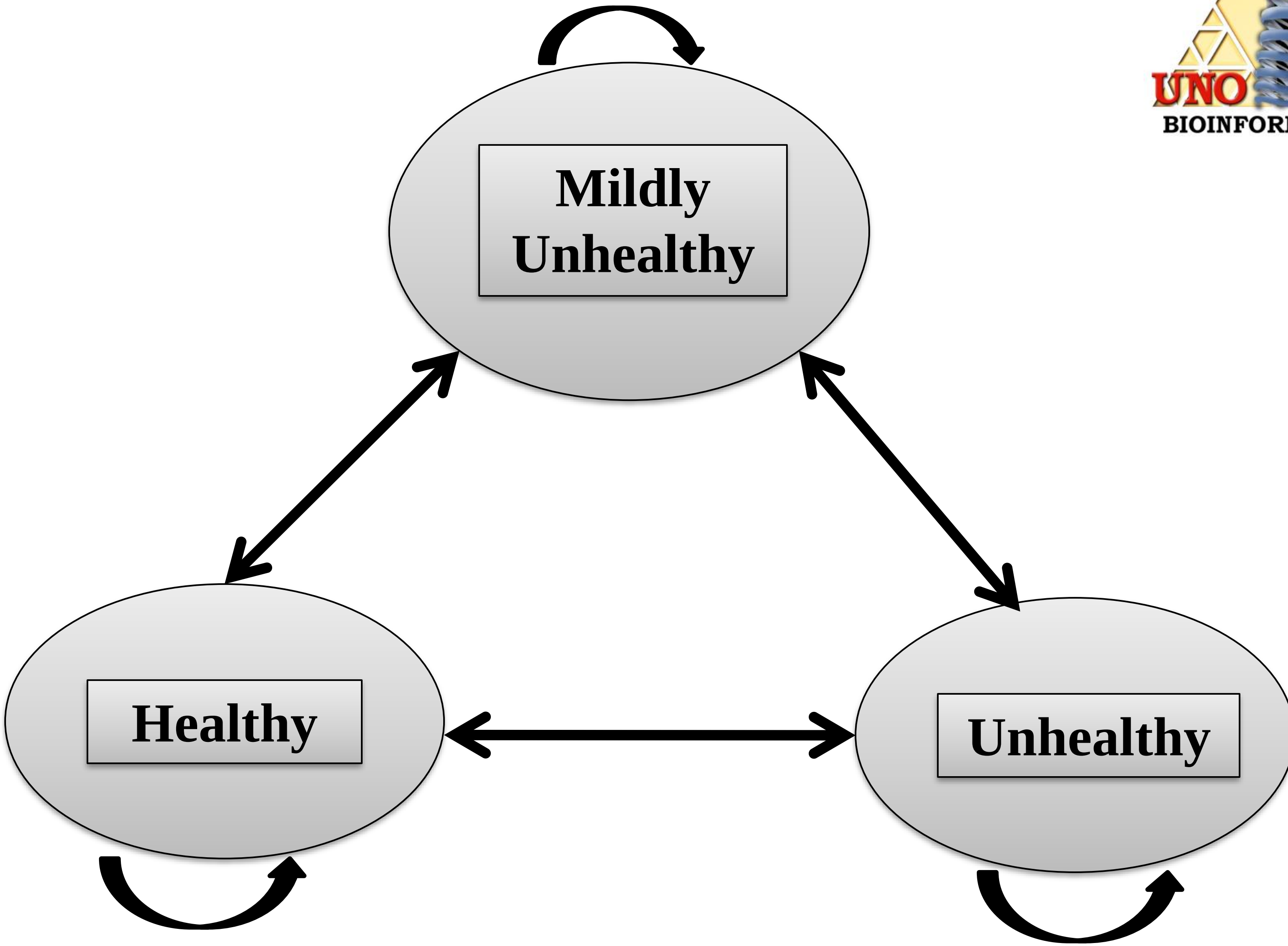
Phase IV: Fall Prediction



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

UNO BIOINFORMATICS



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

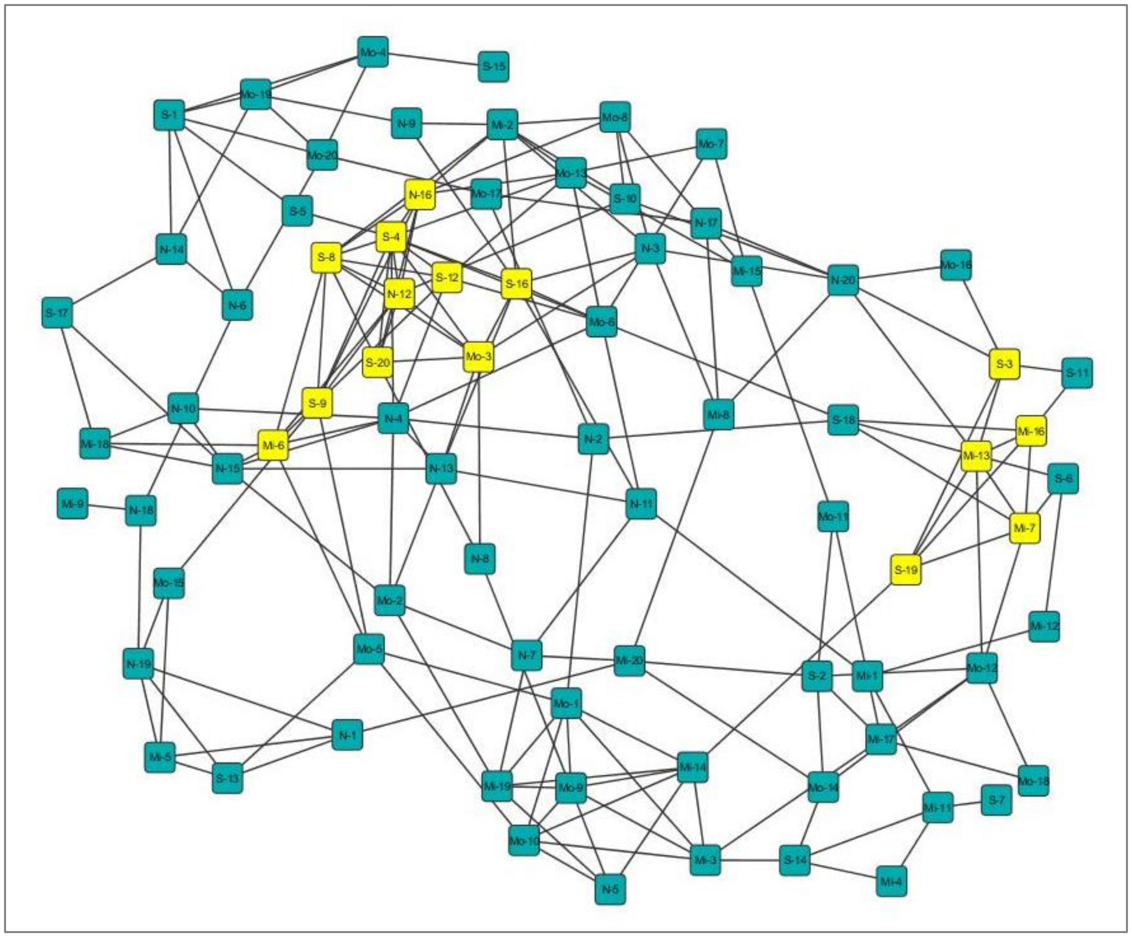
- Scientific Data-Driven Revolution: An Overview
- Big Data Analytics: An Overview
- Heath Monitoring and Data Analytics
- Wireless Sensors and Mobility Analysis for Healthcare
- ***Correlation Networks and Population Analysis***
- Correlation Analysis and Mobility – Network Analysis in Heath Monitoring
- 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 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?

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

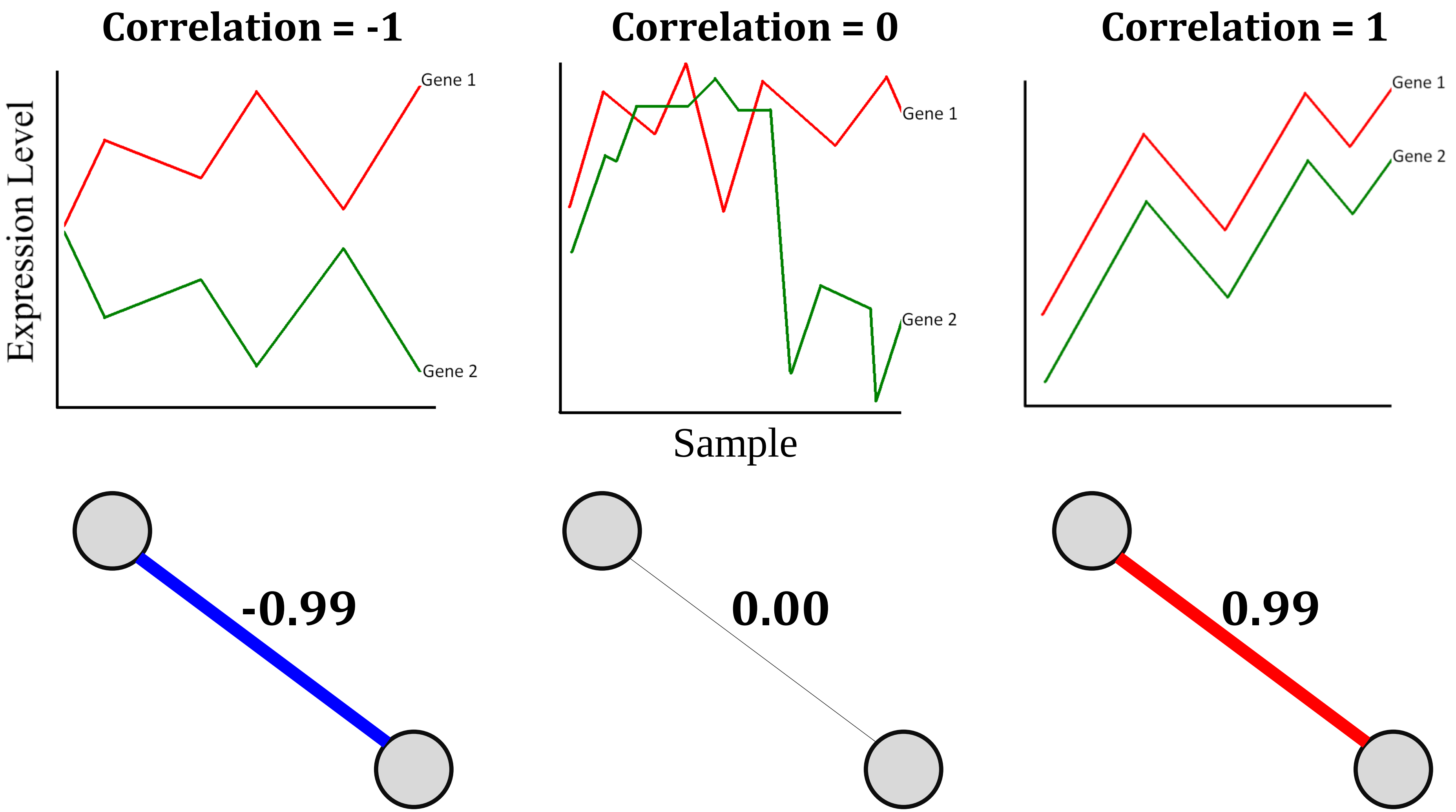
Correlation Network Applications

- "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

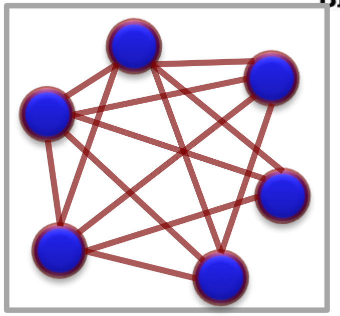
Correlation Network Analysis



Local Network Structures

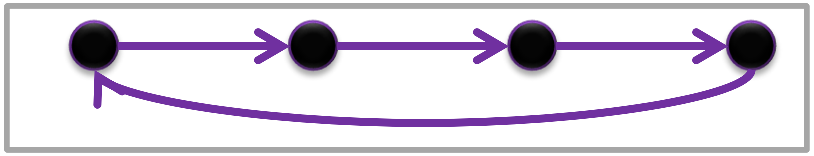
- **Cliques**

  Protein complexes, regulatory modules



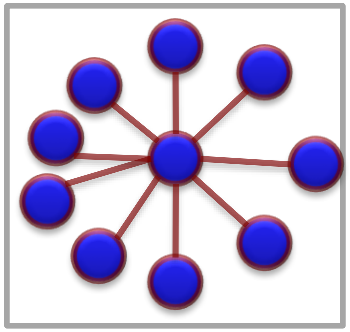
- **Pathways**

  Signaling cascades

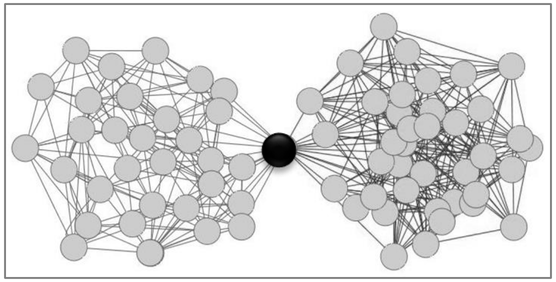


- **Hubs**

  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/fucntion 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

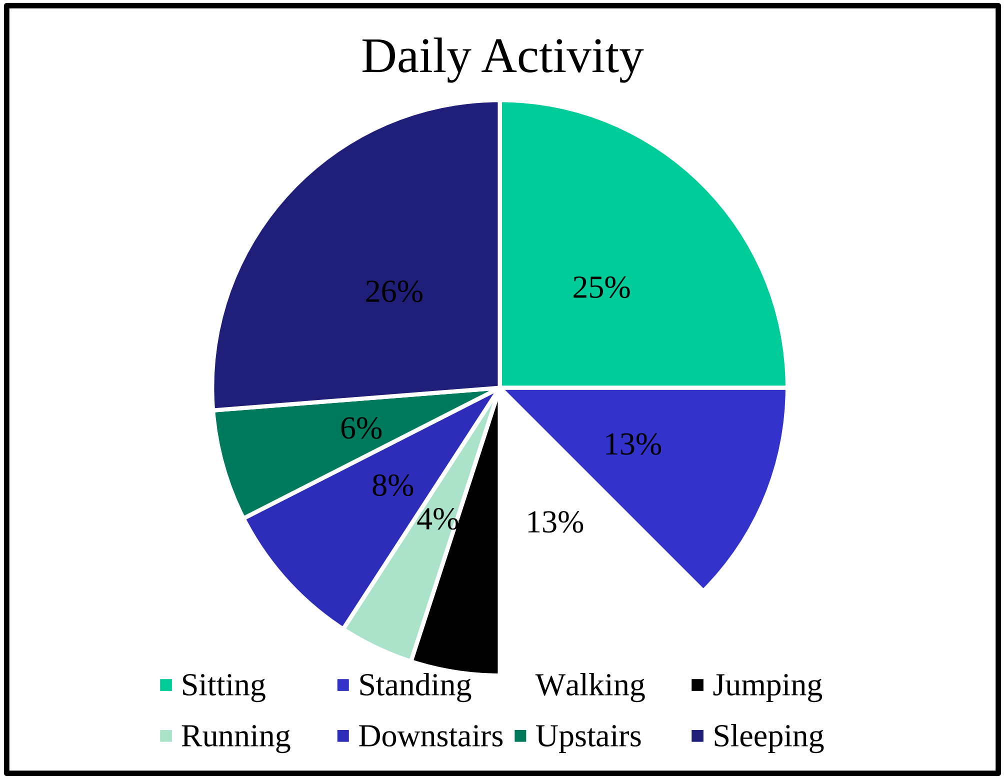
- Mobility parameter derived using the weighted average activities performed and the metabolic (or a unified parameter such as Energy Expenditure) equivalents of these activities.
- Weighted average of the time for which a particular activity is performed is captured, and the metabolic equivalent is used to assign the point to gauge mobility.
- The importance of Context

Mobility Parameters

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



Physical Activity	METS
Walking (2 mph)	2.5
Dancing	2.9
Walking (3 mph)	3.3
Walking (4 mph)	4.5
Jogging (10 min miles)	10.2
Climbing	6.9
Lying	1
Sitting	1
Standing	1.2

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

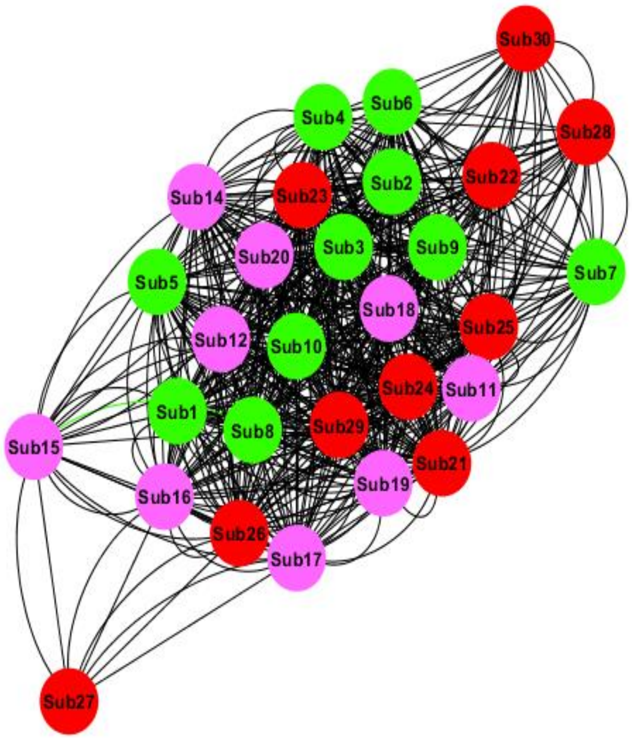
- Sample Generation Setting
  - Weighted activity level value
  - Each group has different mobility decline rate per hour
  - Group1 - 10%/hour, group2 - 20%/hour and group 3 – 30%/hour (shown in different colors)

	Sub1	-	-	Sub30
Work start	553.78	-	-	384.85
2nd hours	498.40	-	-	269.40
4th hours	448.56	-	-	188.58
6th hours	403.71	-	-	132.01
Work end	363.34	-	-	92.40

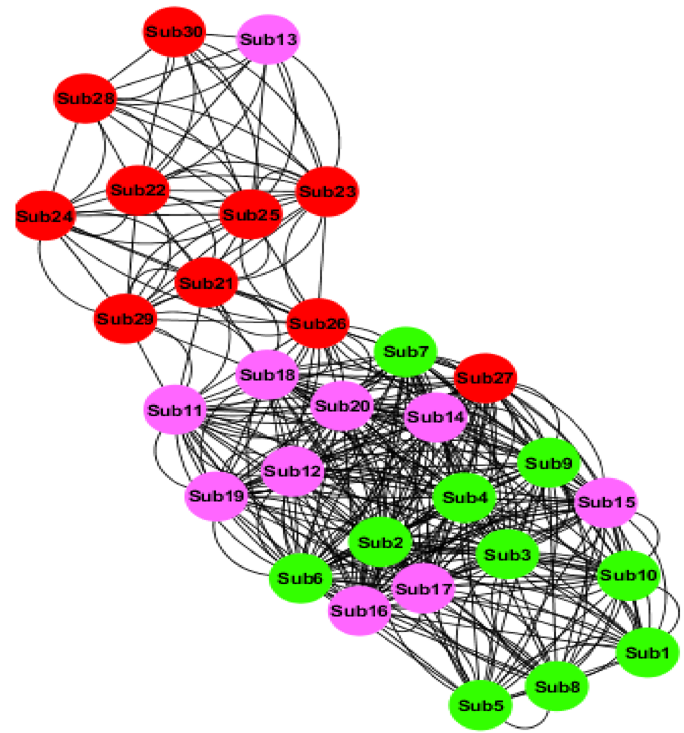
Scenario Description

- 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



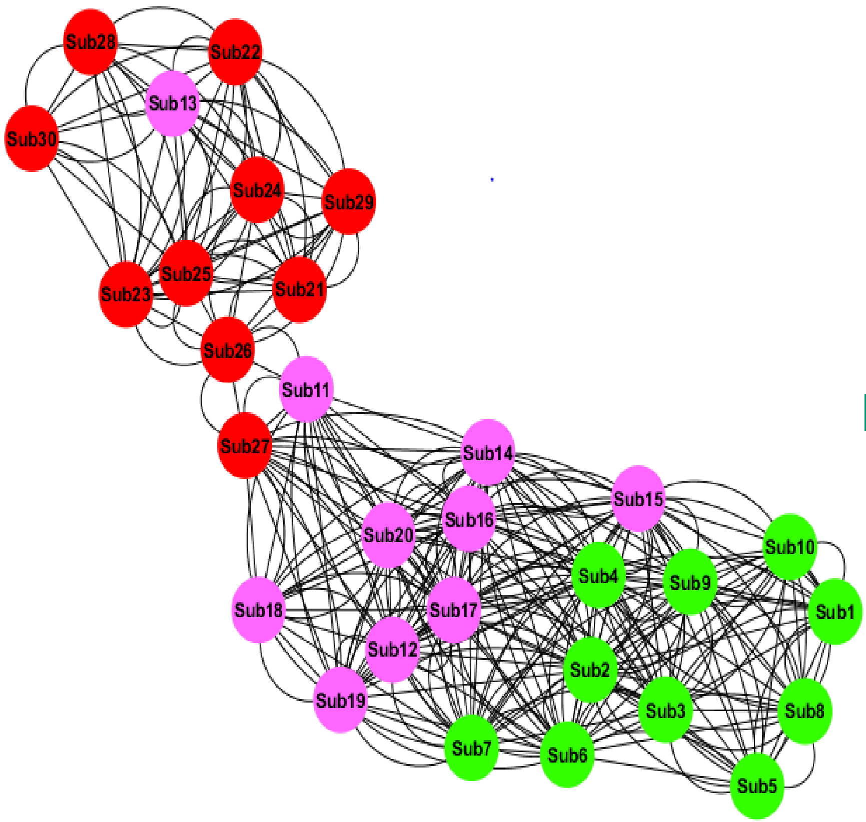
Sample 1



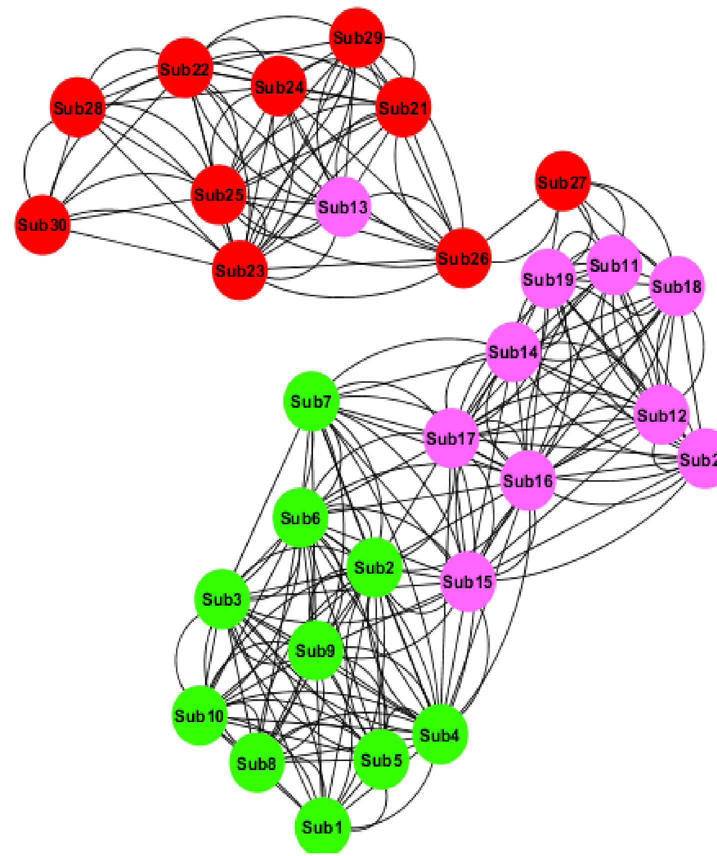
Sample 2

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

Time Series



Sample 3



Sample 4

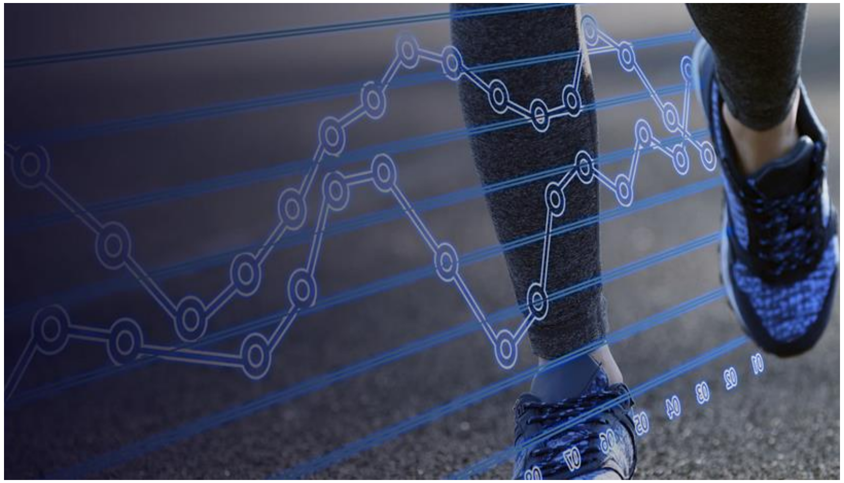
Green nodes – Group 1
Pink nodes – Group 2
Red nodes – Group 3

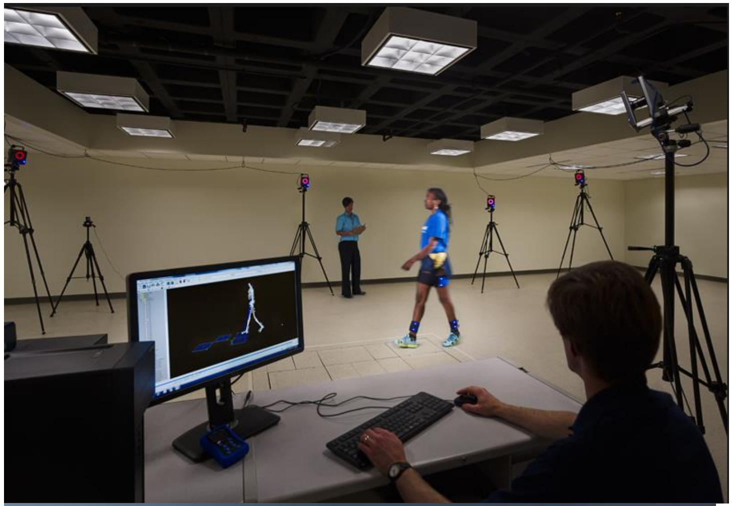
Simulation Results

- 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

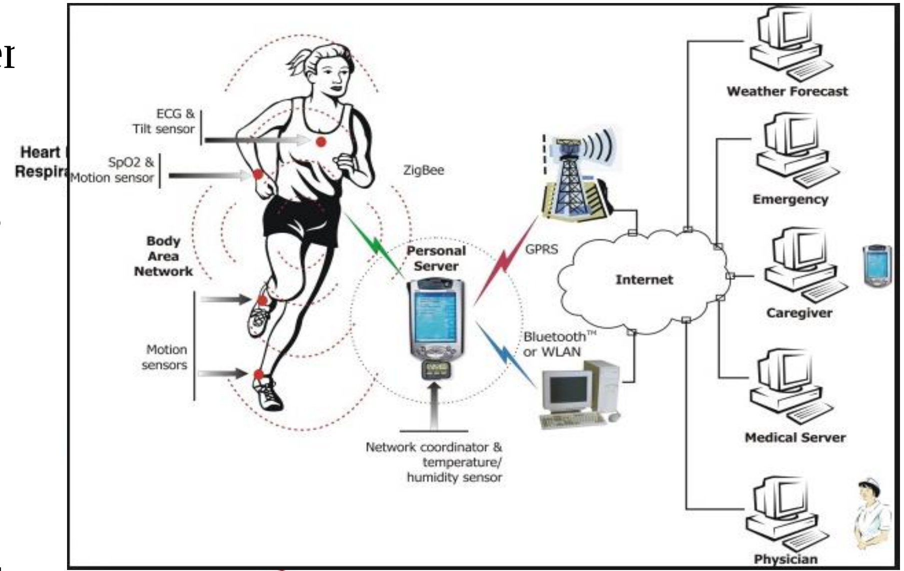
- Gait parameters and health conditions
- Traditional gait measurement systems
- Advanced technologies
- Continuously and remotely monitoring





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



Correlation Network Model

UNO BIOINFORMATICS

- **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

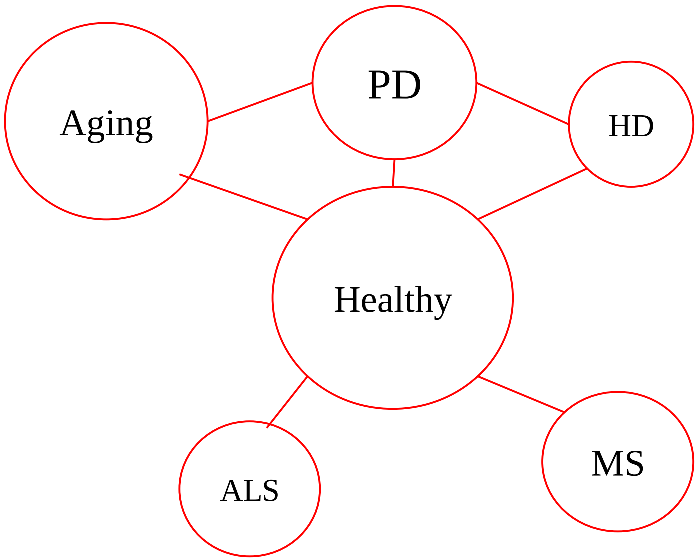
✓ Network Model

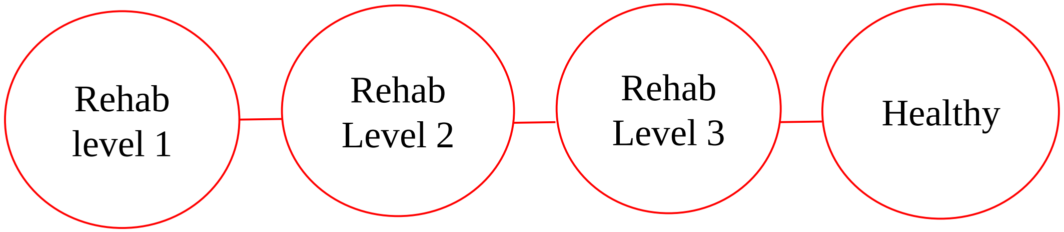
Identifying and Extracting Discriminating Parameters

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

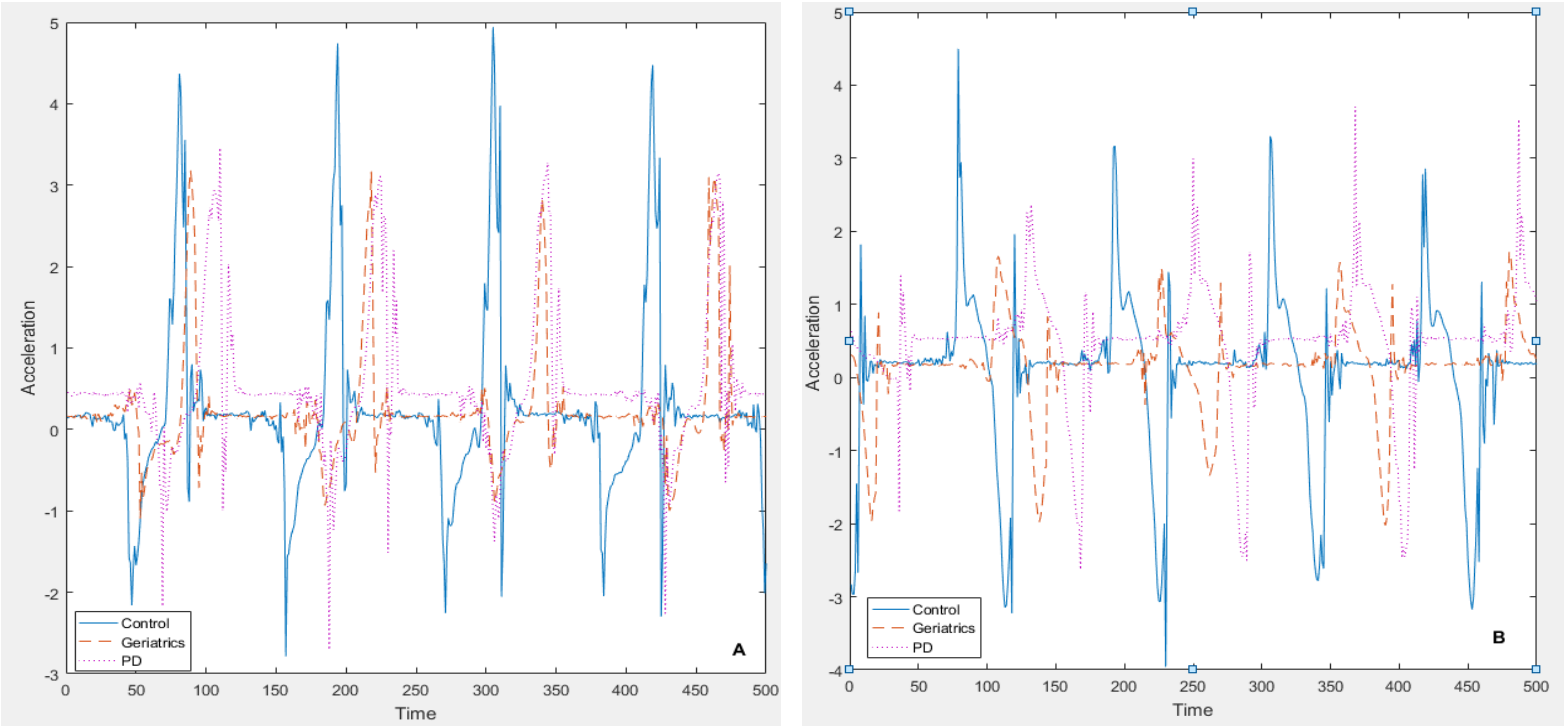
Participants and Protocol

- Participants

	Control	PD	Geriatrics
Number of subjects	5	5	5
Gender (M/F)	3:2	3:2	2:3
Age	64 (10)	72 (6.3)	81 (5.9)
UPDRS III		20.8 (6.1)	
H & Y		2.6 (0.5)	

- 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

$$\text{Intensity} = \sum \quad x^2 + y^2 + z^2$$

- Symmetry

$$\text{Symmetry} = \frac{\text{Min}(Fl, Fr)}{Max\ (Fl, Fr)}$$

Stride Segmentation and Parameter Extraction

Parameter	What this parameter measures	Axis
Avg_stride_Time	**Average stride Time**	-
Acc_Var	Standard deviation of stride_to_stirde acceleration (stride-to-stride acceleration **variability**)	X, Y and Z
CV_StrideTime	Stride time coefficient of **variation**	-
Avg_RMS	Magnitude of acceleration **(Intensity)**	X, Y and Z
Max_Acc_perStride	Mean value of maximum magnitude of acceleration per stride	X, Y and Z
Avg_VecMag	Vector magnitude of acceleration in all directions **(Intensity)**	-
Stride_time_sym	Stride time **symmetry** between left and right leg	-
VecMag_sym	vector magnitude **symmetry** between left and right leg	-

ANOVA and Post Hoc test

UNO BIOINFORMATICS

Parameters	Control vs Geriatrics		Control vs PD		Geriatrics vs PD	
	Mean difference	sig.	Mean difference	sig.	Mean difference	sig.
Avg_stride_time	**-0.23****	**0.004**	0.53	0.621	**o.18***	**0.021**
Var_Acc_X	0.037	0.075	0.41	0.059	0.003	0.960
Var_Acc_Y	0.049	0.063	-0.19	0.603	**-0.68***	**0.011**
Var_Acc_Z	0.077	0.051	0.02	0.681	0.056	0.091
CV_StrideTime	**-0.22****	**0.010**	**-0.28****	**0.009**	0. 43	0.111
Avg_VecMag	**0.04****	**0.004**	-0.21	0.119	**-0.62*****	**0.000**
Avg_RMSX	**0.37*****	**0.000**	0.039	0.728	**-0.34*****	**0.000**
Avg_RMSY	**0.43*****	**0.001**	-0.14	0.278	**-0.58*****	**0.000**
Avg_RMSZ	**0.19***	**0.040**	0.33	0.882	-0.16	0.091
Avg_MaxAcc_X	**1.67*****	**0.000**	**0.72***	**0.020**	**-0.95*****	**0.001**
Avg_MaxAcc_Y	**1.87*****	**0.000**	0.38	0.562	**-1.48****	**0.004**
Avg_MaxAcc_Z	1.01	0.140	-0.46	0.616	**-1.48***	**0.027**
Sym_VecMag	0.42	0.323	**0.83***	**0.030**	0.41	0.341
Sym_StrideTime	0.32	0.080	0.28	0.011	0.012	0.231

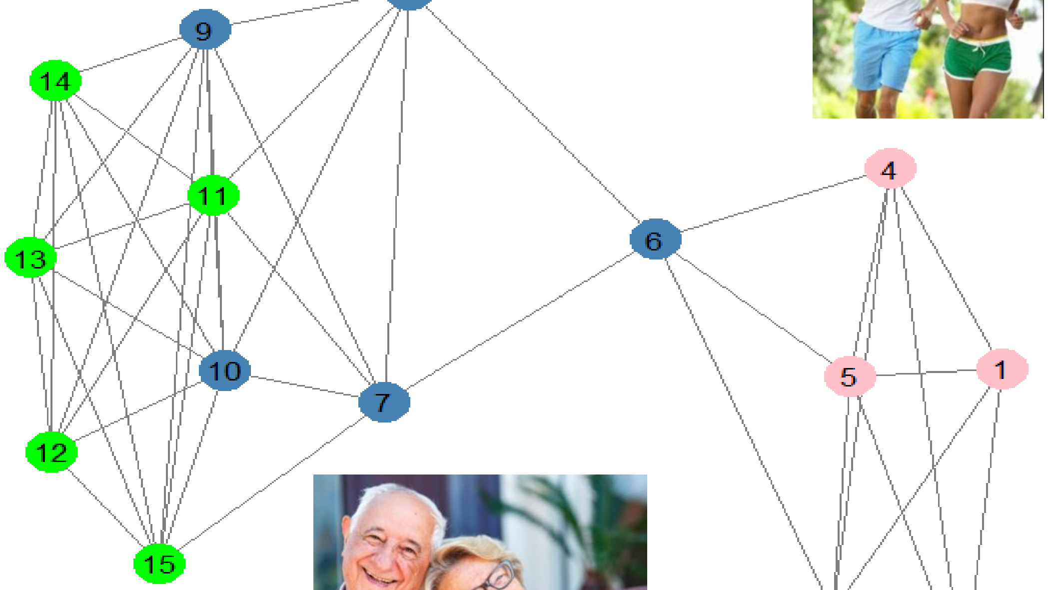
Correlation Analysis and The Network Model

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

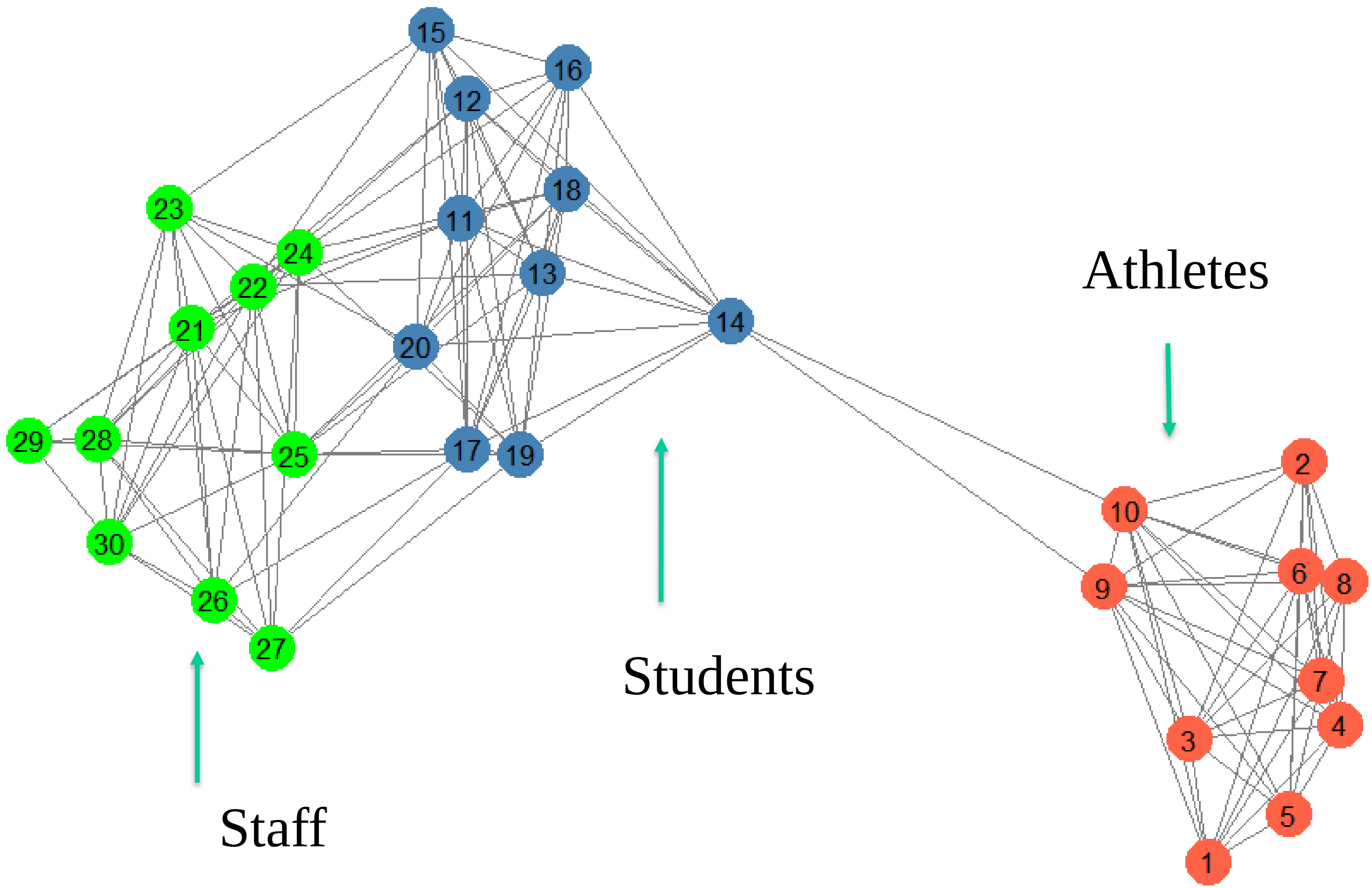
- **Creating Network Model**
  - **Vertices represent subjects**
  - **If two subjects are highly correlated, there is an edge**

The Network Model

- Clusters
- Cliques
- Gateway Vertices
- Transitions



The Dynamic Aspect

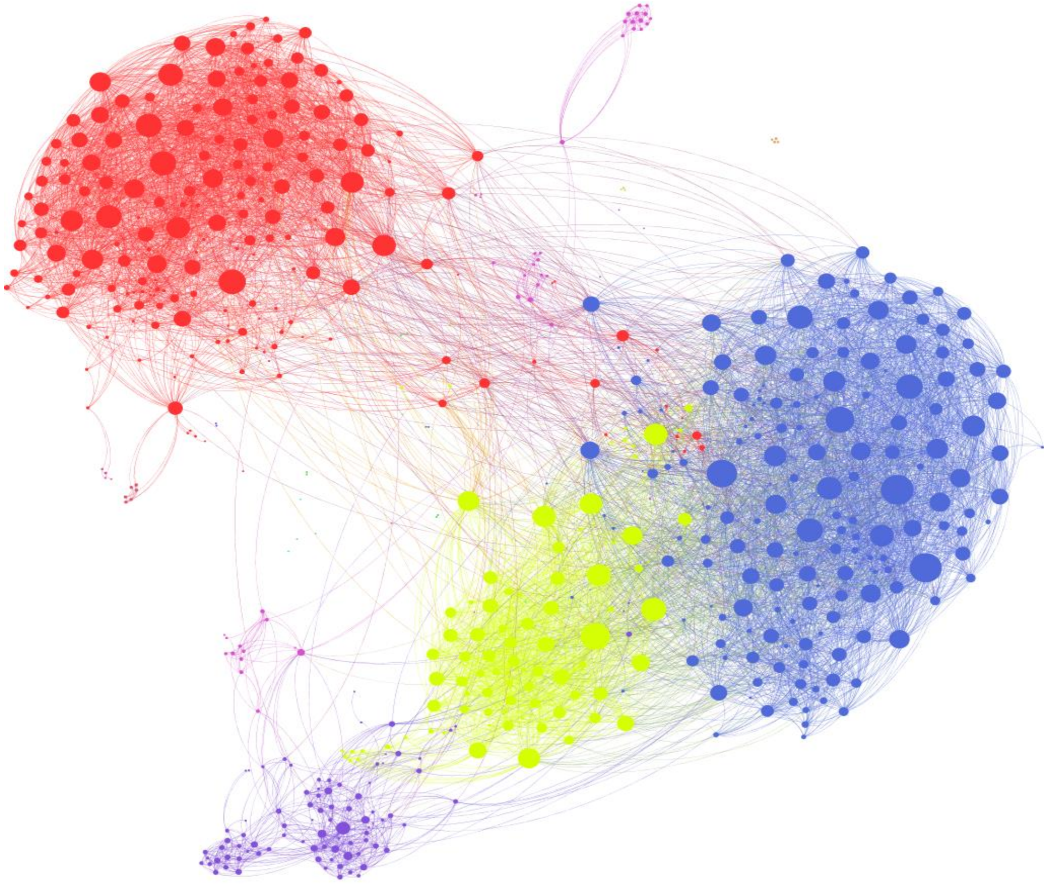


Summary

- 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

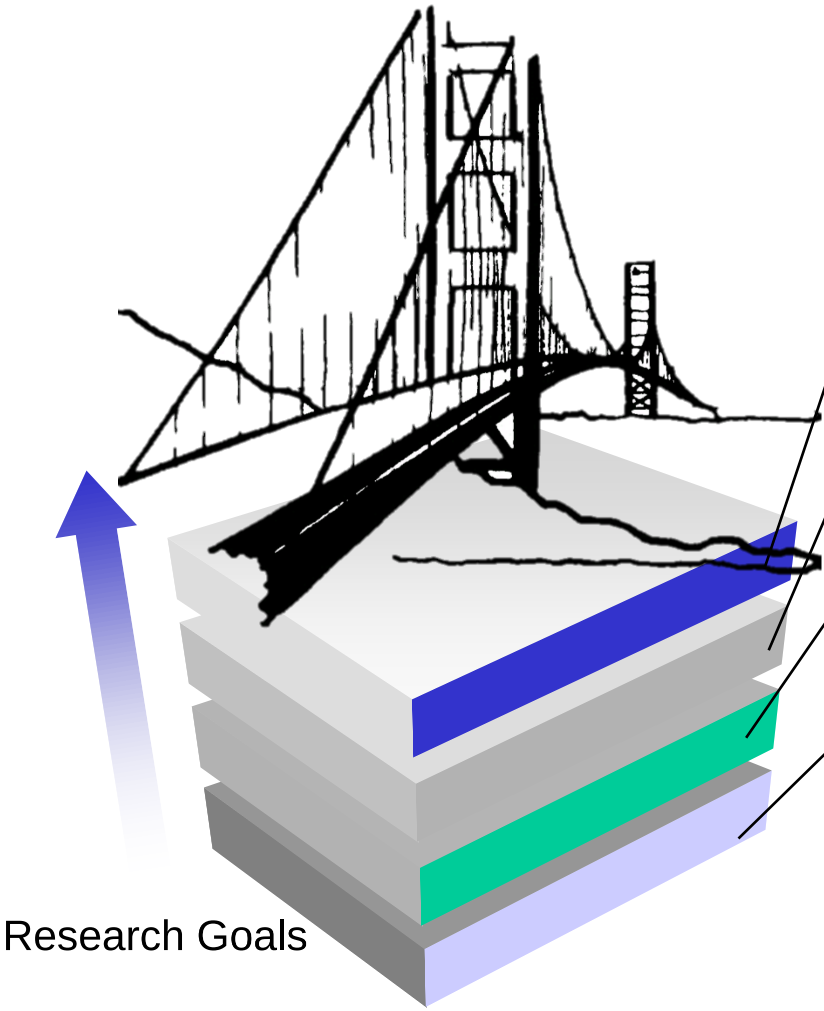
Applications for Patient Care: Patient care after discharge

- 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

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

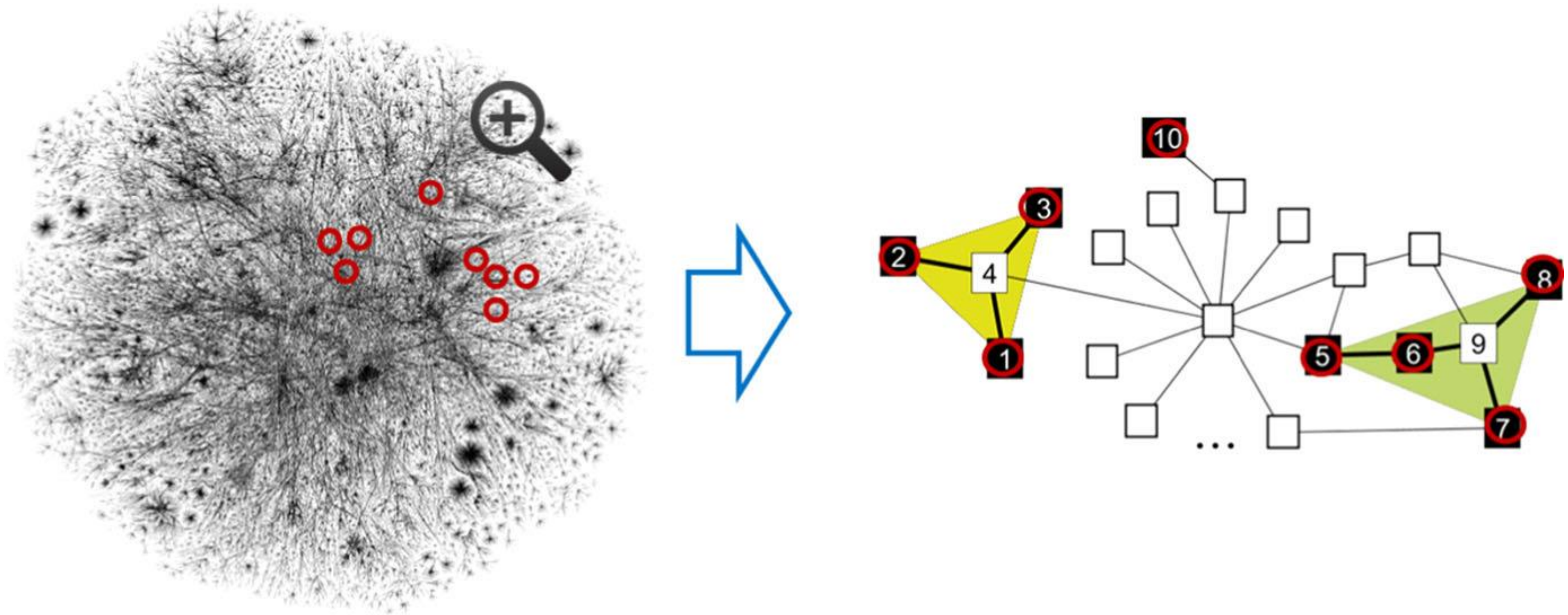
Issues in Structural Health Monitoring

- 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

	Indicator	**Assessed through**
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

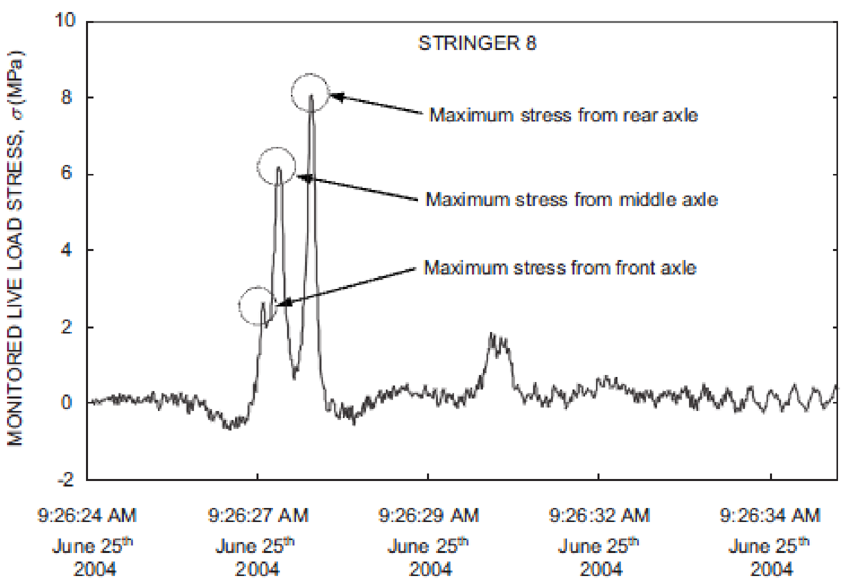
Anomaly detection with graphs

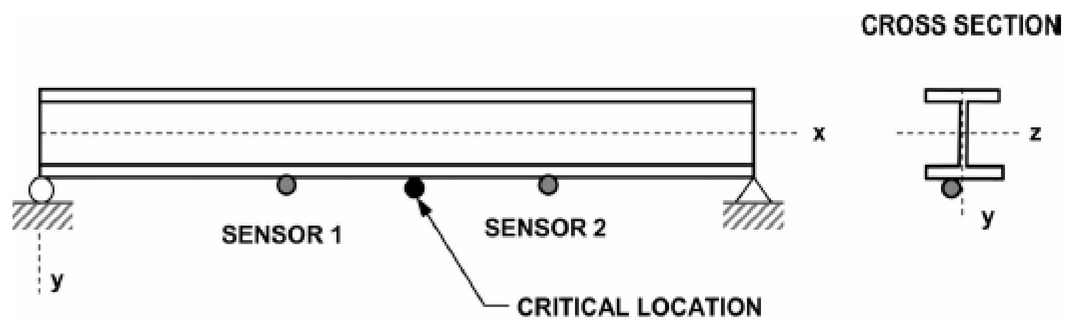


- Interactive graph querying:
  - Left: Given a set of detected abnormal nodes
  - Right: Further grouping into few clusters. Detecting how the abnormal nodes are linked to each allows to find undetected abnormities with similar characteristics

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**,…





	SENSOR 1	SENSOR 2	CRITICAL LOCATION
LOCATION	$s_1(x,y,z)$	$s_2(x,y,z)$	$s_c(x,y,z)$
MOMENT OF INERTIA	$I_{z,1}$	$I_{z,2}$	$I_{z,c}$
BENDING MOMENT	M_1	M_2	M_c

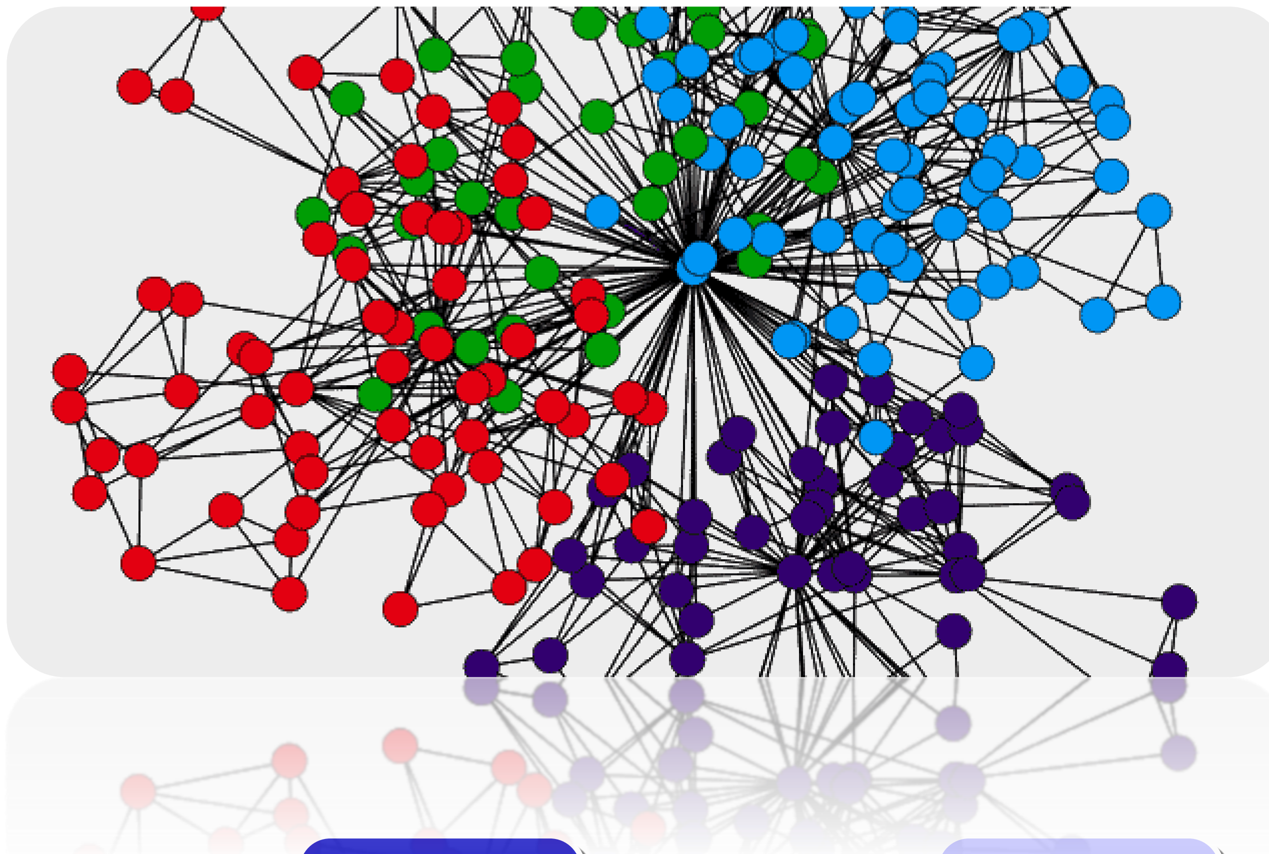
Anomaly Detection Example 2: Railway Track Performance Monitoring

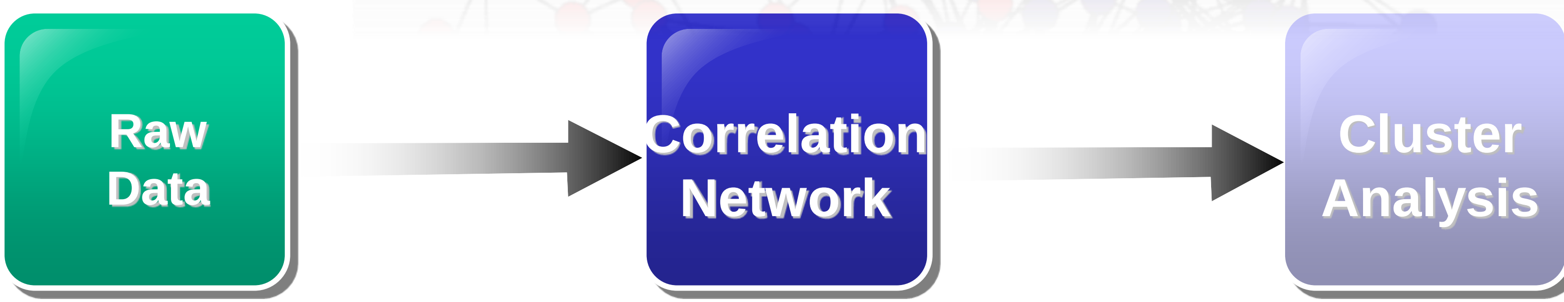
- Addition to previous contribution by thermal stress

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

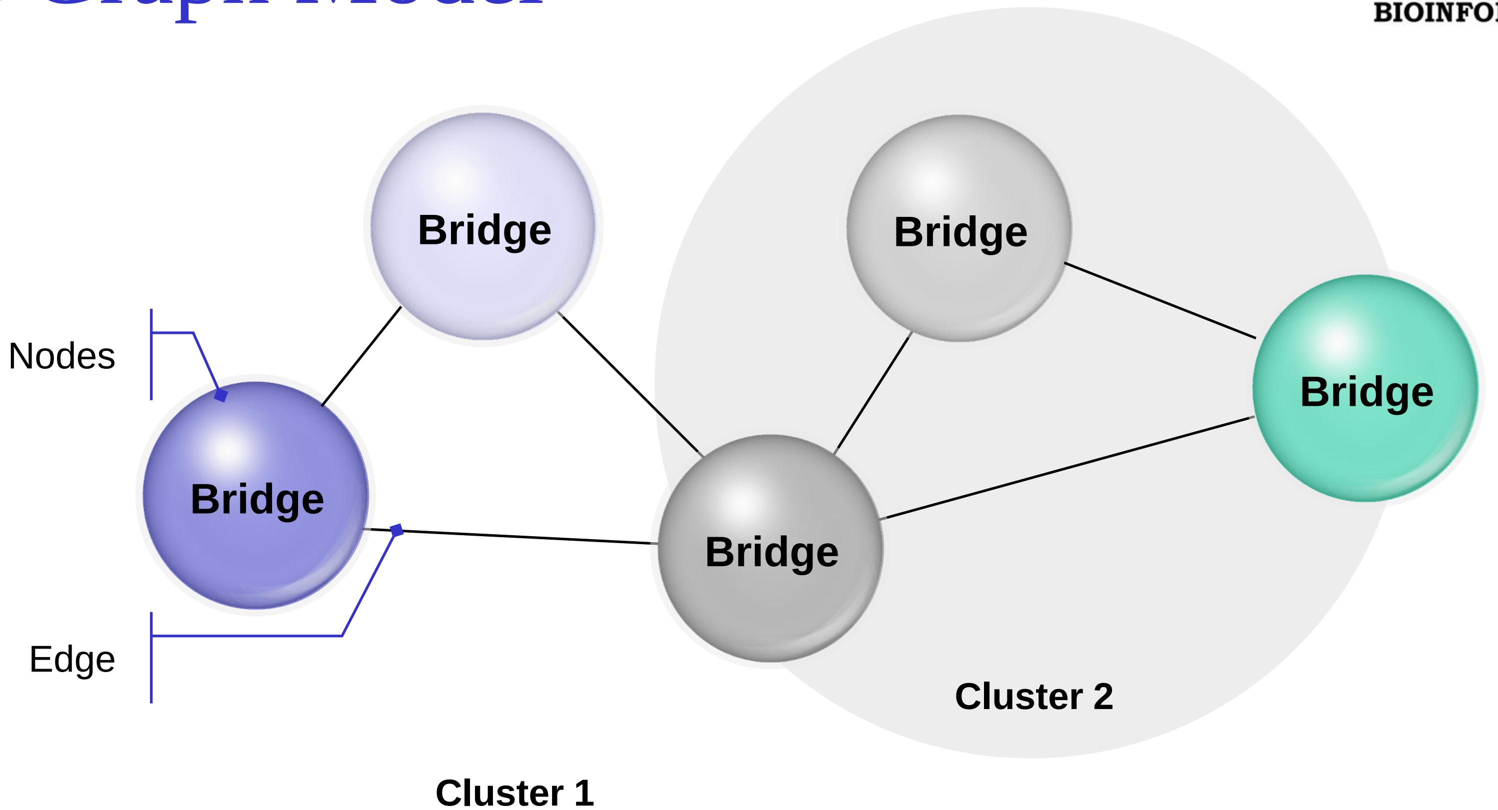
- Relevance for this research
  - Bridge Membrane Monitoring
  - Steel Constructions

Bridge Correlation Networks





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

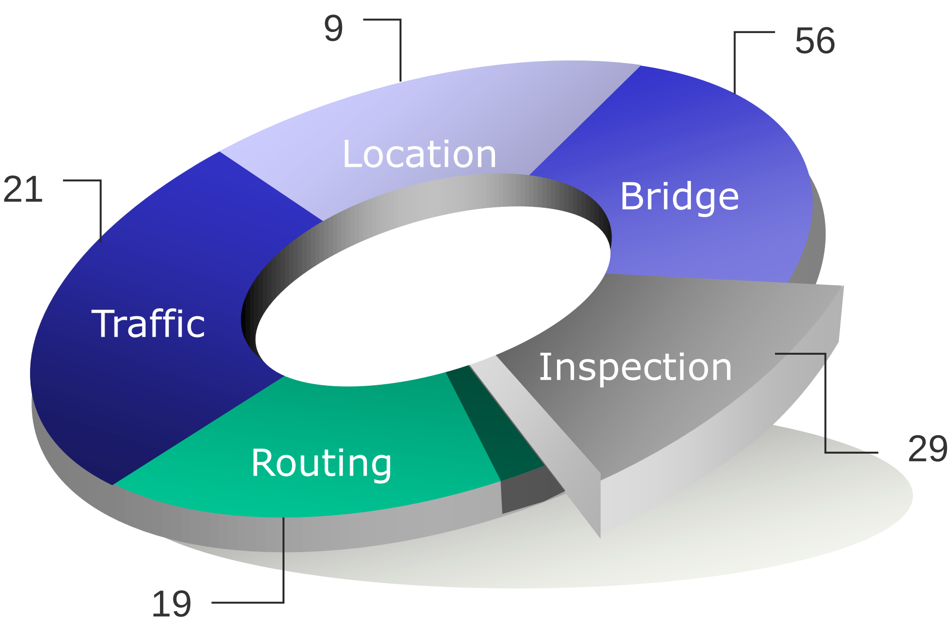
Official NBI Data

http://www.fhwa.dot.gov/bridge/nbi.cfm

Online Database (NBI & Bridgehunter)

http://uglybridges.com

NBI Data Analysis



Inspection

Suff. Rating

Struc. Eval.

Deck Geometry

Deck Con.

Structure Condition

Channel

Bridge

Railings Transitions

Year Built

Year Update

Type of Design / Construction

Clearance

Traffic

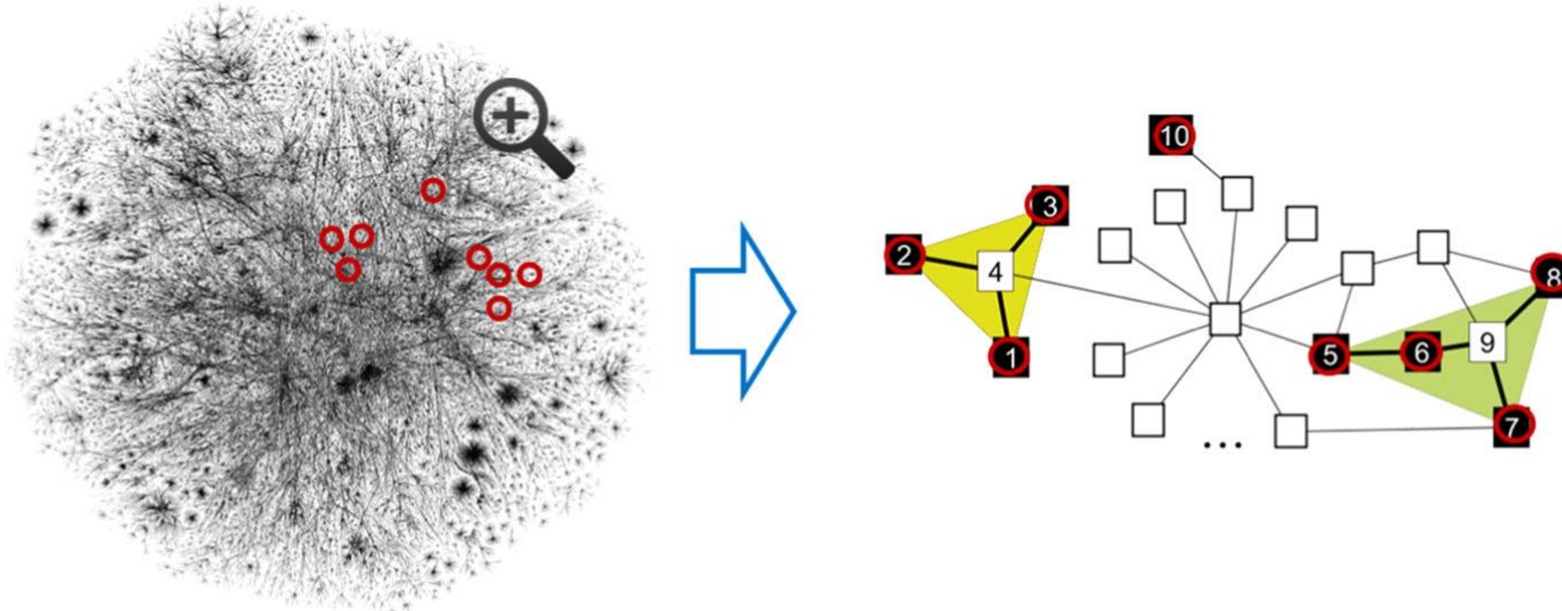
Avg. daily Traffic

% Truck Traffic

Lanes

Highway?

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 abnormities with similar characteristics

Dynamic Graph Models

… change through

- Sensor inputs
- additional data sources
- data manipulation

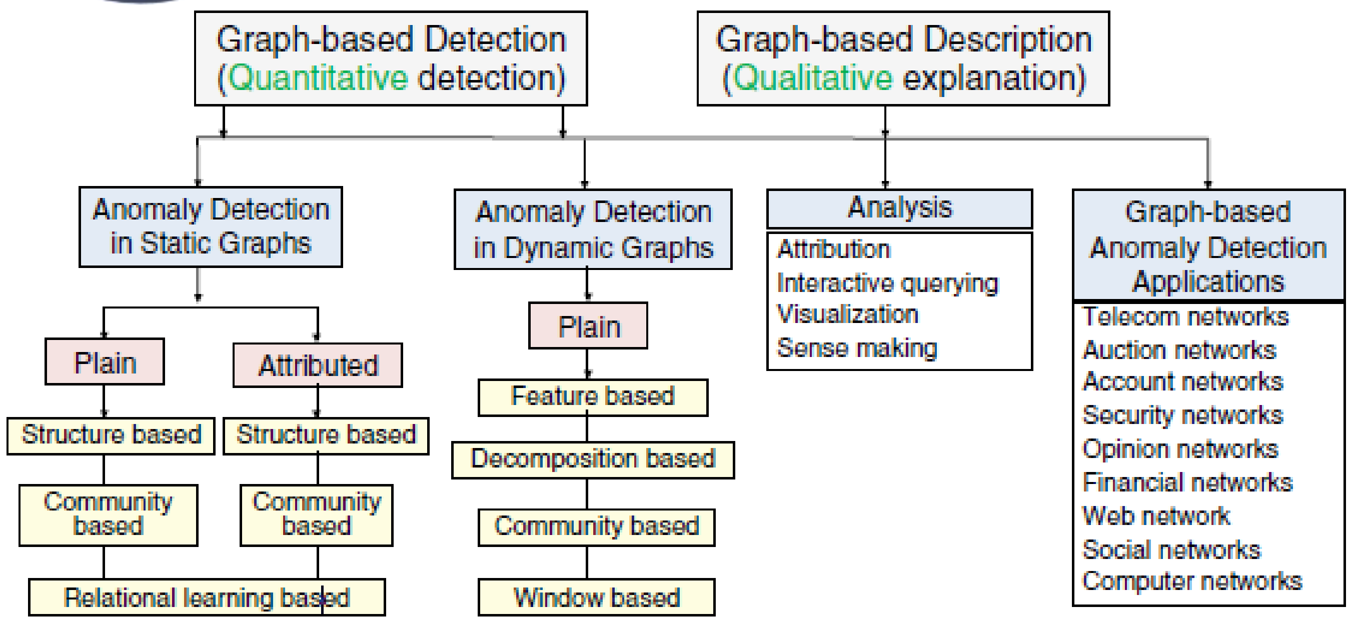
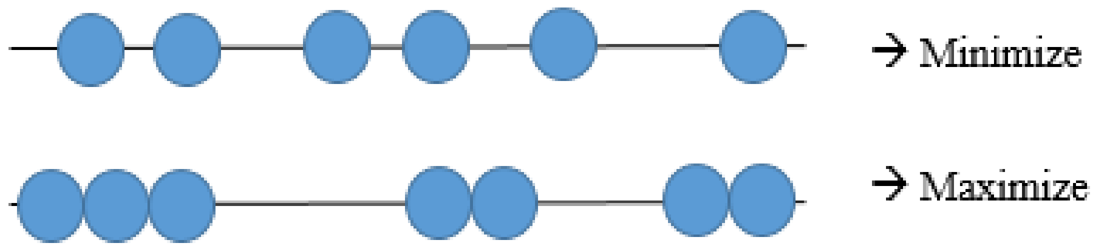


Fig. 2 Graph-anomaly detection: the outline of the survey

Source: Akoglu, Leman, Hanghang Tong, and Danai Koutra. "Graph based anomaly detection and description: a survey." Data Mining and Knowledge Discovery 29.3 (2014): 626-688.

Clustering / Prediction Model

$$sim_{a,b} = \sum_{1\ldots n}^{i} c_i * v_i$$



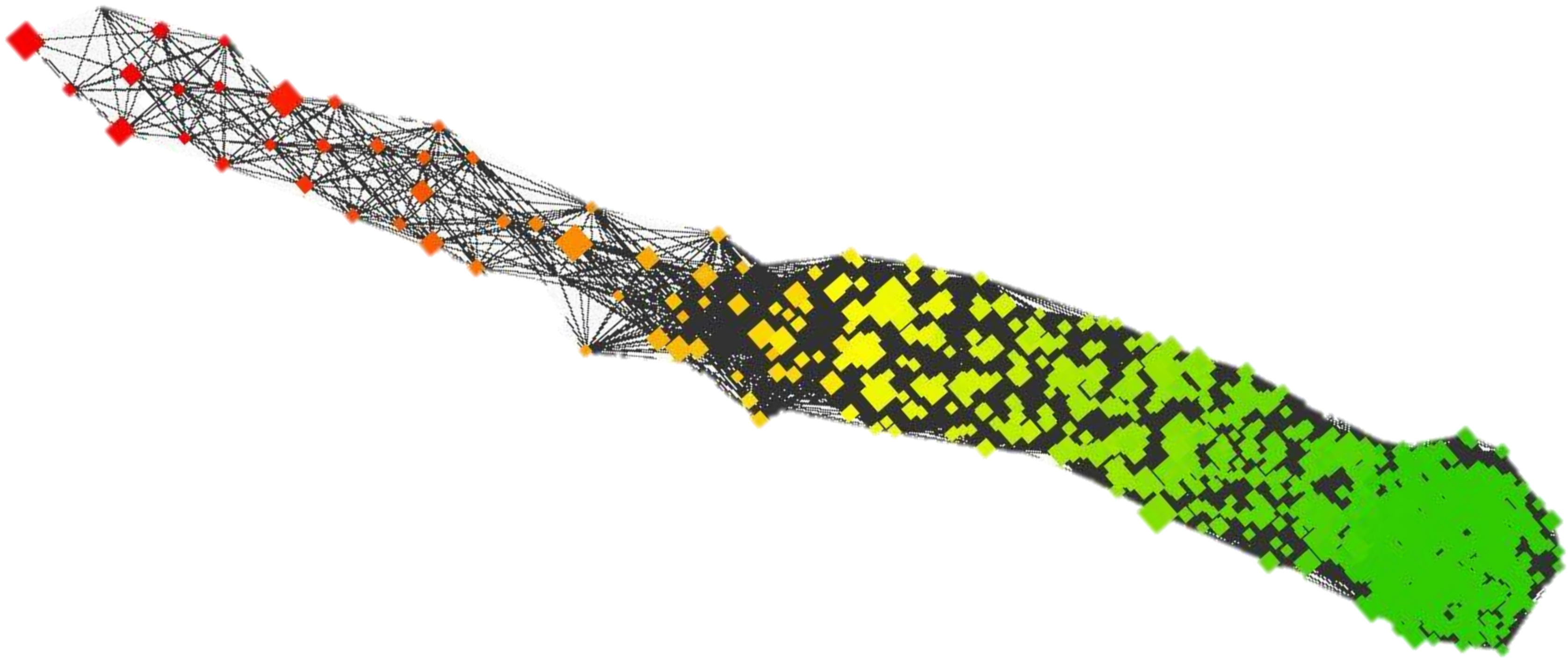
A
Generic

Generic Approach
no data preparation
Can find unknown problems

B
Specific

Specific Approach
faster learning
more in-deep analysis

Visualization Age / Safety



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

Correlations based on:

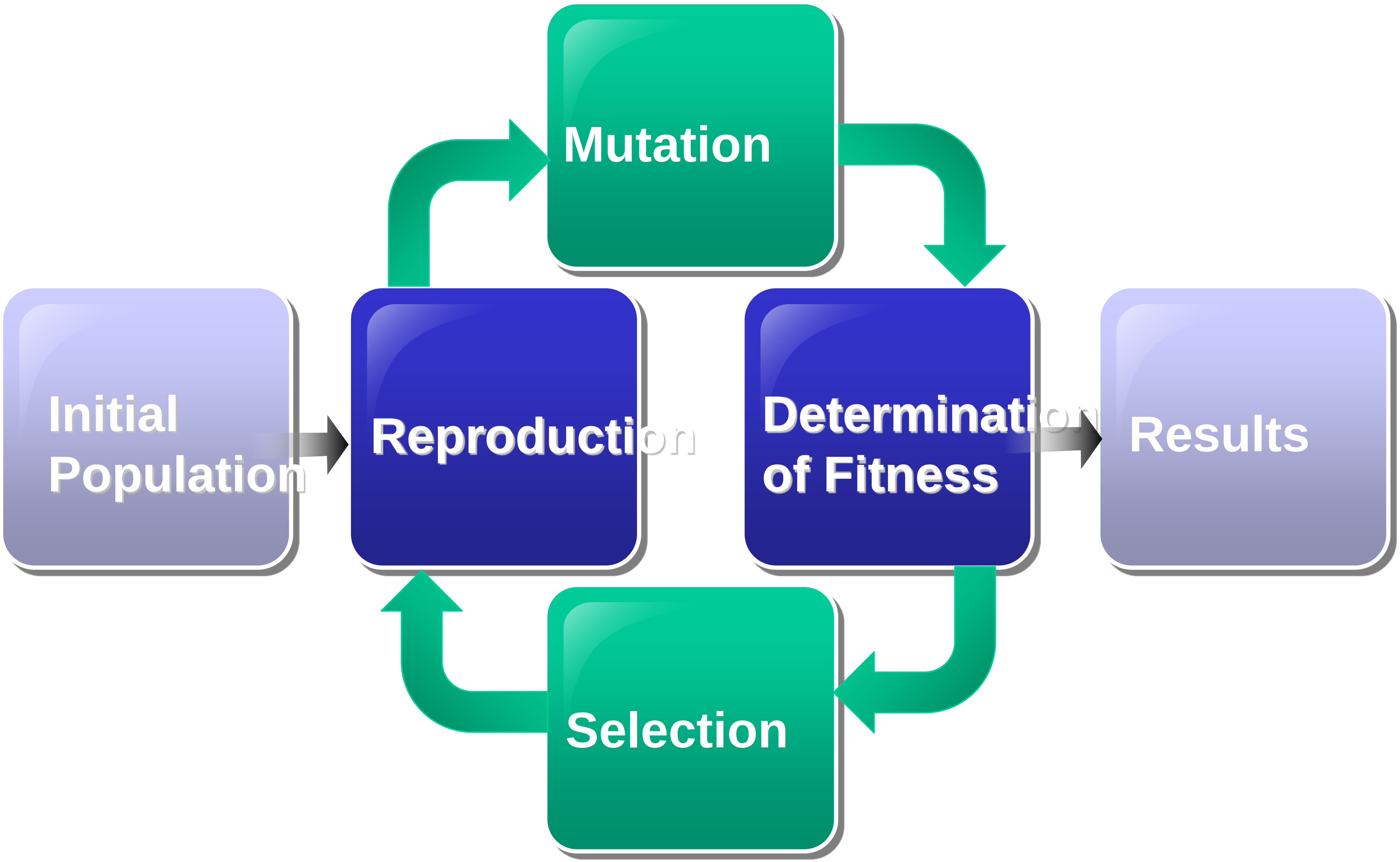
Age

SR

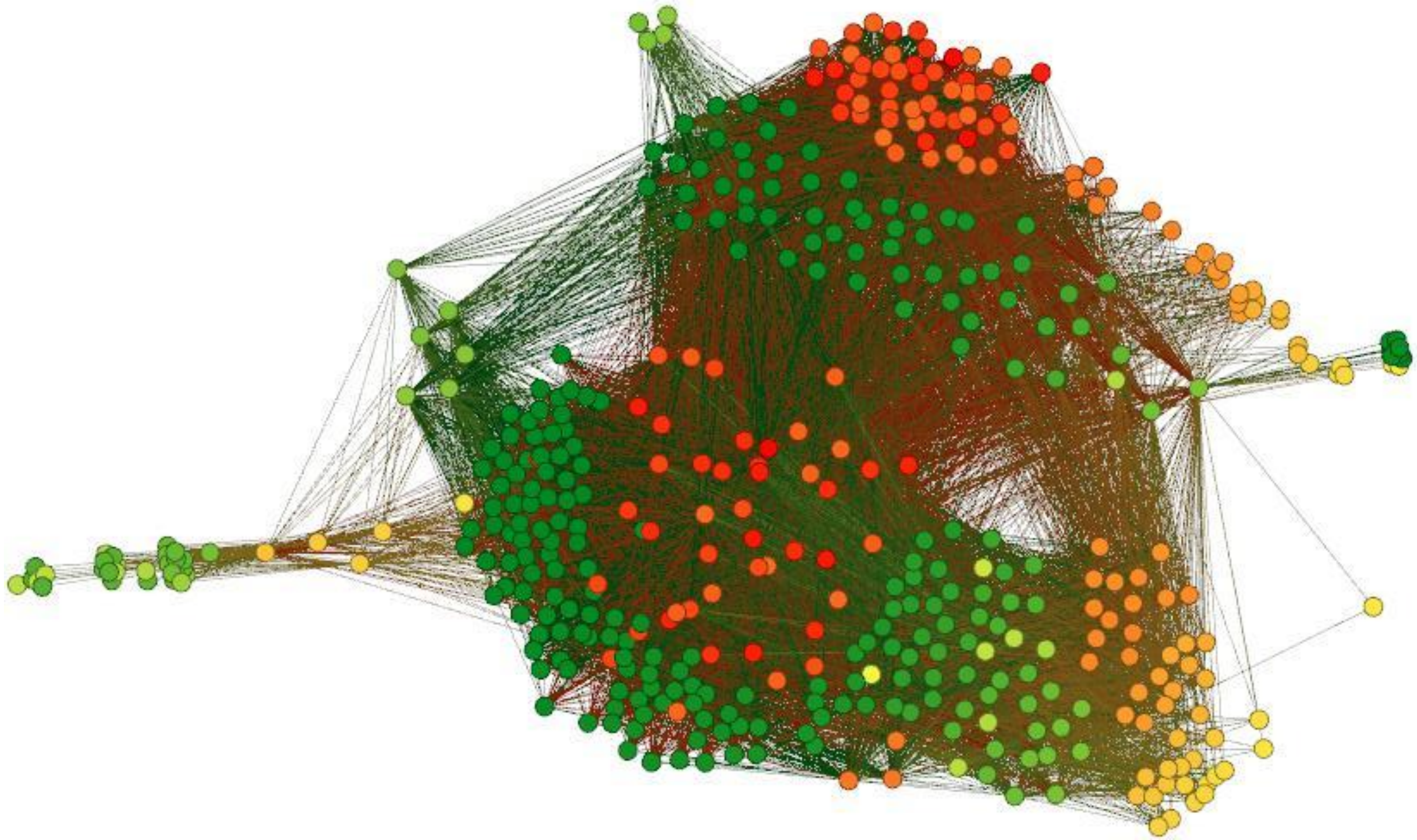
Coloring based on:

SR

Prediction: Genetic Algorithms



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:

SR **CON**

Coloring based on:

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

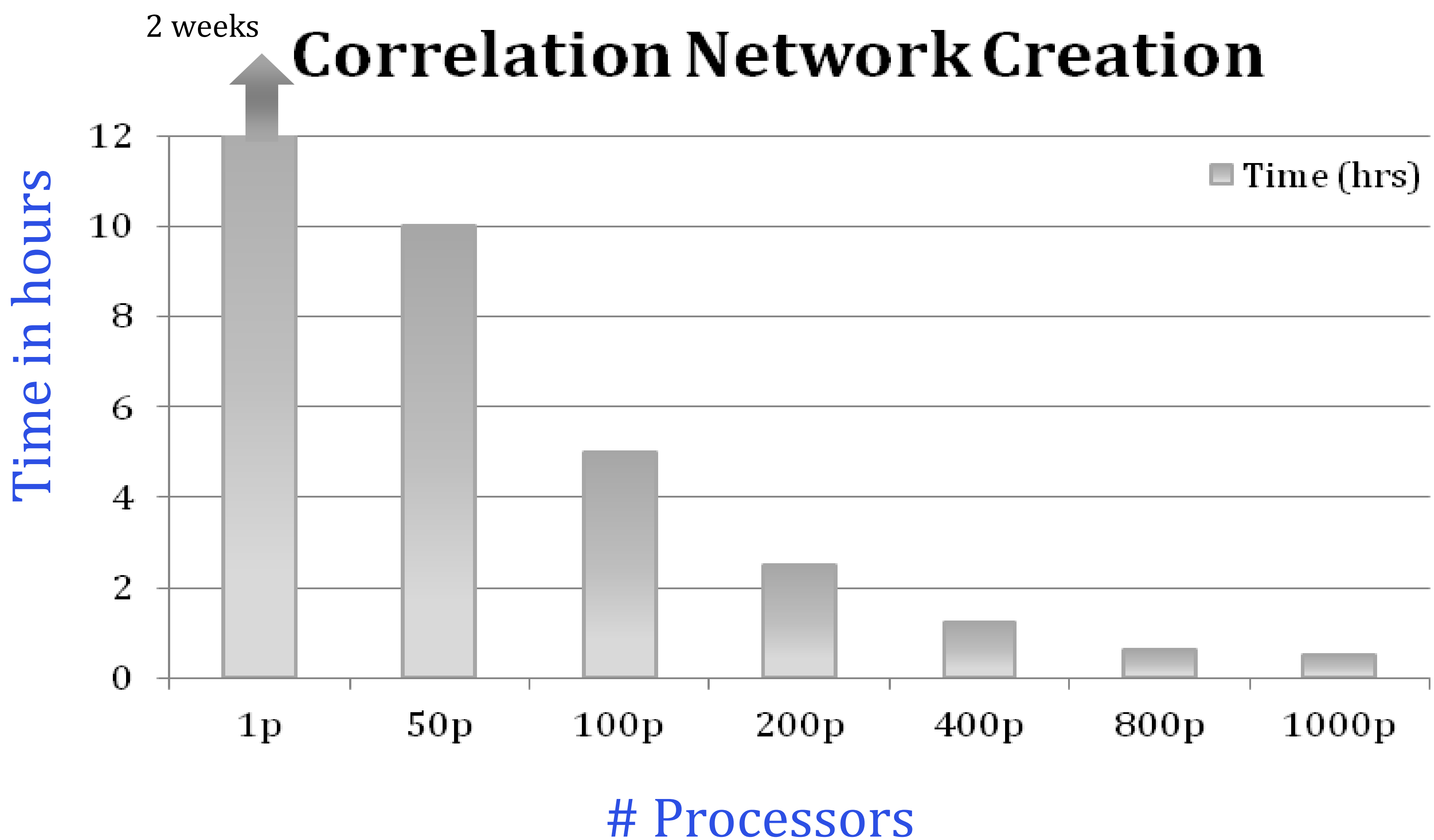
HPC and Big Data

- 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



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

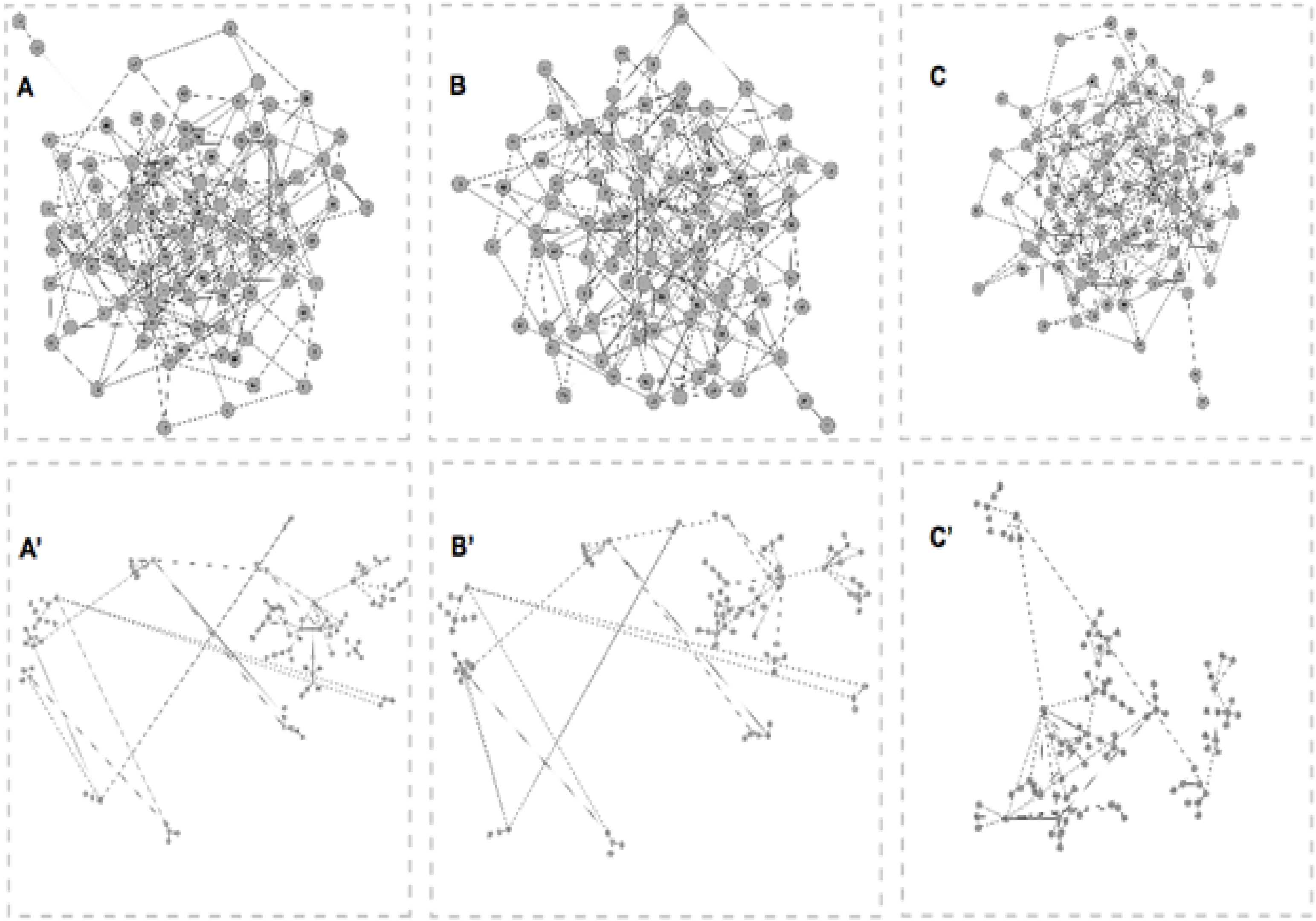
- Chordal graph sampling
    - Keep triangles in expression graphs
    - Remove large cycles, extra edges
    - Keep clusters, identify new clusters
- Spanning tree sampling
    - Keep high degree nodes (maybe?)
    - Remove up to 50% of edges
    - Enhance identification of lethal nodes
- Planar Graph 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 w
    - ➢ 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|*max_deg)**

Need to Maintain Key Structures

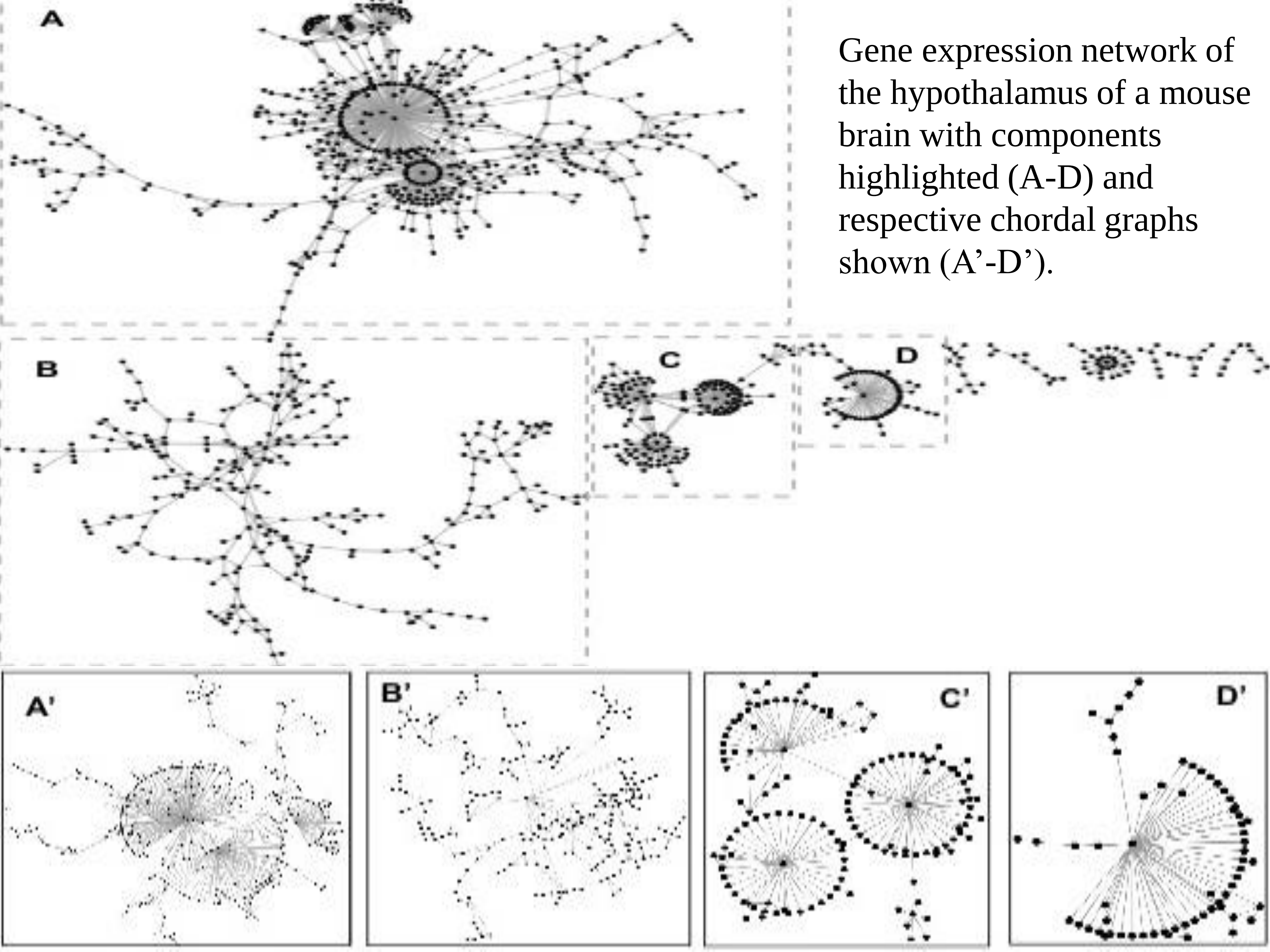


Proposed Approach

- ✓ 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

Network	Conserved clusters	Newly identified clusters
Young Mouse RCM Ordering	1	6
Young Mouse BFS Ordering	1	3
Middle Aged Mouse BFS Ordering	4	7
Middle Aged Mouse RCM Ordering	4	4

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