18-755 Networks in the Real World (F'17)

Suggested Projects

Tentative Projects for 18-755
- Fall 2017-

I. Smart Cities and Connected Communities

Project I.1	Phone call record data and economic prosperity (1-2 students) [Mentor: Kartikeya Bhardwaj]
Overview	Can you determine the economic prosperity in a region using call data records? Can we estimate poverty maps from phone call networks? Typically, economic prosperity can be measured in terms of the number of economic opportunities one gets. Phone call networks can act as a proxy of how diverse a person's interactions are within a society. The more diverse the interactions, the more chances of getting new opportunities! In this project, we will explore…
Tasks	Direction 1 (for a 2-person team): ● Gather call data records (telecommunications datasets) for Italy. For instance, certain regions in Italy (e.g., Milan and province of Trentino) have datasets ranging from social networks, news pulse data, to weather and energy data, all the way to telecommunications data, available. Also, collect some economic indicator data (e.g., income data, household budget, etc.), either directly from country-specific census or other sources (e.g., create an approximate proxy of economic indicators using nightlight satellite data from NOAA). This will give a way to analyze correlations between telecommunications and economy in a region. ● Construct phone call networks from the available telecommunications data. Analyze the properties of these networks (e.g., degree, communities, centrality, etc.) ● Are there any correlations between the telecommunications networks and economic prosperity (e.g., income, household budget data, poverty indices, etc.) for any given region? Are there any correlations between extreme events (e.g., extreme events as indicated by the available news pulse data, or extreme temperatures and/or precipitation) and phone call patterns? Can you characterize people mobility from the phone call networks? Direction 2 (for a 1-person team: without the economic indicator part): ● The same dataset in Italy also has other data such as social networks, weather, and energy networks, news pulse, etc. Can we see any correlations between social pulse, news pulse, and telecommunications networks? How are telecommunications affected by weather, etc.?
References	[1] Eagle, Nathan, Michael Macy, and Rob Claxton. "Network diversity and economic development." Science 328.5981 (2010): 1029-1031. [2] Barlacchi, Gianni, et al. "A multi-source dataset of urban life in the city of Milan and the Province of Trentino." Scientific data 2 (2015): 150055. [3] Mellander, Charlotta, et al. "Night-time light data: A good proxy measure for economic activity?" PloS one 10.10 (2015): e0139779. [4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4405276/pdf/pone.0124160.pdf [5] https://www.nature.com/articles/srep04001.pdf

Project I.2	Citibike Usage Balancing through Congestion Pricing (2 students) [Mentor: Kartikeya Bhardwaj]
Overview	Citibike is the New York City bicycle sharing scheme, and is the largest such system in the USA. Bicycles are locked in stations distributed all throughout the city. Users make one-way trips by unlocking a bike from one of the departure station docks, riding to a station close to their destination, and locking the bike into a dock at that station. The current pricing is based on long- and short-term subscriptions, typically linked to a credit card. Short trips (e.g., under 35 min) are included in the price of the subscription, while longer rentals are progressively penalized by additional cost, in an attempt to prevent users from monopolizing bikes. The system suffers from one major drawback: users tend to follow unidirectional flows at various times of the day. For example, in the morning everyone tends to ride away from the train station to their

	workplace. As a consequence, it is often the case that there are no bikes in the morning near the train stations. The opposite situation also occurs quite often. The user may find all docks occupied at the destination and cannot return the bicycle. Instead, they must find another station with empty docks, often quite far away, at the cost of extra time and frustration. The existence of an app that accurately indicates the location and state of every station in the city, offers only meager redemption. Currently, Citibike attempts to rebalance the system by moving bikes in bulk by truck or van from full stations to empty ones but the results are less than spectacular. This project is about improving this situation via optimization and network modeling.
Tasks	● Introduce a congestion-based pricing scheme. For instance, an user should pay more for unlocking a bike from a nearly empty station and for locking a bike at an almost full station. Conversely, an user should be paid to unlock a bike from an almost full station and for leaving it at a nearly empty one. Ideally the scheme should be cost neutral, simple to deploy, and easy to explain to the general public. ● Make use of the enormous amount of Open Data available from nyc.gov. This includes data from bicycles, pedestrians, bike stations, citibike station status feeds, traffic, etc. Construct various kinds of networks – citibike network, pedestrian mobility networks, traffic networks, etc. using the NYC open data. Do you notice any similarities between these networks? ● Build weighted networks of trips for each month and o Perform community analyses at multiple scales (e.g., neighborhood, borough, city) o Compare network structure and dynamics across weeks, months, etc. Are there any seasonality effects? ● Feel free to formulate other/additional analyses; interpret and visualize your findings
References	● http://citibikenyc.com/system-data ● Agent-based simulation' environment: https://en.wikipedia.org/wiki/NetLogo ● http://www.treehugger.com/bikes/citi-bike-reveals-how-new-yorkers-ride-their-bikes.html Citibike mobile app for iOS and Android gives a good idea about the system

Project I.3	**Criminality Networks (2 students) [Mentor: Fillipe Condessa]**
Overview	How does criminality in Shadyside and Squirrel Hill affect criminality in Oakland? Can Network Science help the Pittsburgh Bureau of Police to better understand the networks of crime? Is the spatial and temporal propagation of crime the same for all types of crimes? This project explores the network nature of crime in urban scenarios (e.g., Pittsburgh, Chicago, New York, ...) and its characteristics and dynamics across different types of crime, and in space and time.
Tasks	● Create a dynamic multi-layer (i.e, involving multiple types of crime) network from borough crime activity. ● Create dynamic a geo-spatial multi-layer (i.e, multiple types of crime) network from crime report data. ● Analyze dynamic network properties to identify types of crime, locations, and times that are central on the spread of criminality. ● Identify key strategies based on dynamic network properties for deterrence, detection, and mitigation of crime in cities.
References	[1] http://csce.ucmss.com/books/LFS/CSREA2017/IKE3704.pdf [2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4361400/pdf/pone.0119309.pdf Sample data: https://opendata.cityofnewyork.us/

II. Machine Learning and Network Science

Project II.1	Deep Learning and Satellite Data for Poverty Prediction (2 students) [Mentor: Kartikeya Bhardwaj]
Overview	Do you know that poverty can be estimated with high accuracy using nightlight satellite data? Rather than relying on (paper or phone) surveys which are tedious, time consuming and inaccurate, recent advanced data mining techniques like Deep Learning (with Convolutional Neural Networks) can be used to transfer knowledge from image processing to economic predictions. Explore more on poverty estimation in this project!
Tasks	● Collect all the necessary data listed in references (code and data collection instructions are very well documented for this project). This would include the survey data as well as the nightlight satellite data. Learn to use the publicly available code for poverty prediction. ● Try to introduce a network angle to the problem. For instance, can we discover some similarities between the poverty in different regions using satellite data (e.g., similarities between the features learned by the deep learning algorithm)? If so, then we will obtain a network of poverty across the globe. ● Analyze the properties of this network (degree, communities, centrality, etc.). Predict poverty for regions not considered in the original study. Can we use deep learning trained on certain regions to reliably predict poverty in completely different test regions? ● Can we also use the trained deep learning model to predict something else other than poverty (e.g., health, happiness, social participation, etc.)?
References	[1] Jean, Neal, et al. "Combining satellite imagery and machine learning to predict poverty." Science 353.6301 (2016): 790-794. [2] Xie, Michael, et al. "Transfer learning from deep features for remote sensing and poverty mapping." arXiv preprint arXiv:1510.00098 (2015). [3] http://sustain.stanford.edu/predicting-poverty

Project II.2	From Social Networks to Sustainable Networks: A Machine Learning Perspective (2 students) [Mentor: Kartikeya Bhardwaj]
Overview	Did you know that there are social networking websites dedicated to just sustainability and other causes of high-societal impact? These websites are more popularly known as Green Social Networks (e.g. change.org, carbonrally.com, makemesustainable.com etc.). In this project, we will explore how green social networking can lead to changes in pro-social human behavior towards sustainability and more.
Tasks	● Use Change.org API to get User information (petitions they have created and signed). Create a network of users according to the common petitions available. Also record the geographic locations of the users. ● Next, get information on the petitions (e.g., overview and petition texts, Status: closed, open, victory etc.) and create a network of petitions using Latent Semantic Analysis (LSA, packages available for this in R, Python etc.) and latent semantic similarity analysis. This means that the petitions with high semantic similarity will be linked together. Now try to see if some petitions encouraged some new users to create or sign somewhat similar petitions. The above steps should help answer the following questions: ● How does the dynamics of sustainable pro-social behavior emerge and spread through the network? ● How do the biggest sustainable communities grow/change over time, (ii) Existence/absence of hubs, degree distributions, how hub locations change over time, (iii) Can green social networking platforms actually induce a population level sustainable behavior, etc. ● Come up with other cool questions to answer besides the above. Other tasks include visualizations, etc.
References	[1] D. Yim and S. Viswanathan (2010). "Networks of Green People": Dynamic Network Closure and Prosocial Behavior in Online Communities. Conference on Information Systems and Technology (CIST 2010).

	[2] J. Farell. Network structure and influence of the climate change counter-movement. Nature Climate Change, Nov. 2015. [3] Refer to change.org.

Project II.3	**Inferring Gene Networks (2 students) [Mentor: Chieh Lo]**
Overview	The problem of estimating gene expression networks has attracted a large amount of attention in both biology and machine learning over the past several years. We can model a gene as a variable and the pairwise interactions among them as edges in a graph. In this project, the goal is to first develop a method that estimates these interaction networks and then relate it to external information such as body mass index (BMI),, blood pressure, etc.
Tasks	● Collect gene data including gene functions, signaling pathways, protein-protein interactions, etc. Build a multivariate Gaussian graphical model and learn a sparse estimate of the inverse covariance matrix (also called the precision matrix or concentration matrix) of the distribution [1]. ● Develop graph learning methods and apply it on both synthetic and real-world datasets, and evaluate its performance against the baseline model (Gaussian model). ● Cluster network of genes into different modules and relate it to external information such as BMI. You should be able to answer: what network properties does the gene network exhibit? What are the important genes in terms of network metrics? Can you relate the external information and the gene modules? Report and interpret your findings. ● Required: Basic familiarity with R and Python
References	[1] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. The Annals of Statistics, 2008. [2] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, and M. West. Sparse graphical models for exploring gene expression data. Journal of Multivariate Analysis, 2004. [3] The Gene Ontology, http://www.geneontology.org/ [4] WGCNA: https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/

Project II.4	**Multilayer Microbial Networks (2 students) [Mentor: Chieh Lo]**
Overview	Microbes have huge implications in health and medicine. However, the way in which microbes affect the human health remains largely unknown. Analysis of the human microbiome may help us better understand our own genome. In this project, you are given a dataset that contains the abundance of different bacteria species at different sites in the human body (e.g., gut). The main goal is to discover the function and species interaction relatedness by utilizing multi-layer network algorithms.
Tasks	● Collect sequence data (e.g., 16S rRNA) [1] and generate synthetic data. Use existing methods to build the functional and interaction network (e.g., [2]) and analyze it using network metrics (e.g., degree distribution). ● Utilize multi-layer network algorithms to find the functional and species relatedness. You should be able to answer the following questions: what network properties does the functional and species interaction network exhibit? How many distinct communities exist at each site? Can you find some interesting patterns among these two networks? ● Come up with other interesting questions to answer besides the above. Other tasks include visualizations, interpretations of the community structure. ● Required: Basic familiarity with R and Python
References	[1] HMP: http://hmpdacc.org/overview/about.php [2] Inferring Correlation Networks from Genomic Survey Data. Friedman J, Alm EJ (2012) Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol 8(9): e1002687. [3] Sparse and Compositionally Robust Inference of Microbial Ecological Networks. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, et al. (2015) Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLoS Comput Biol 11(5): e1004226

Project II.5	Smart Campus: Learning Student Movement Behaviors (2 student) [Mentor: Ryan Kim]
Overview	Other than class periods, are there significant patterns of student movement on campus? And why do they happen? Can we predict where students will be concentrated throughout the day? How do campus events influence these behaviors? Students can belong to many communities (e.g., department, clubs, extra-curricular activities, etc.). Can we detect these communities and then discern interesting behaviors (e.g., ECE students spend much less time at the UC than the normal student)? This project will quantify and characterize how students move throughout the CMU Pittsburgh campus. In particular, this project will focus on the detection of rare and unique behaviors, and the relationship between events and behaviors.
Tasks	● Collect datasets: Using a set of Bluetooth-enabled devices, non-invasively collect Bluetooth signals across CMU Pittsburgh campus. Collect a list of events (including advertising methods) that may have influenced student behavior on campus. ● Define a few questions you find interesting in this space. Some possible themes are: 1. Nearly everyone on campus has a department affiliation. Using the datasets, can we detect and classify Bluetooth devices into these departments? Is there interesting information about a certain department community? 2. How do events influence student behavior? How does the advertising (social media, email, posters, fliers) of these events affect its ability to influence behavior? What types of events are the most likely to create large changes in behavior?
References	[1] https://arxiv.org/ftp/arxiv/papers/1702/1702.00546.pdf [2] M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, pp. 479-482, 2008. doi: 10.1038/nature06958 [3] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A Tale of Many Cities: Universal Patterns in Human Urban Mobility," PLoS ONE, vol. 7, no. 5, 2012. doi:10.1371/journal.pone.0037027

III. Knowledge Discovery and Social Networks

Project III.1	Social Network Reconstruction from Partial Information (2 students) [Mentor: Fillipe Condessa]
Overview	Is anonymity persistent across different social networks? How to unmask a secret criminal network, or protect a secret group from discovery from an Orwellian government? The inference of unmeasurable data from structural properties grants us the ability to look beyond the observable domain. This project will explore the application and extension of different inverse problem techniques from signal processing and machine learning to network science to reconstruct partially obscured real subnetworks (with focus on Reddit analysis).
Tasks	● Construct a static multilayer network from Reddit data including multiple subreddits (multiple layers). ● Apply and extend signal processing / image processing techniques for network reconstruction. ● Learn inter and intra layer parameters and priors for sub-reddit membership for users based on their ego-network. ● Reconstruct user membership (sub-reddit) from partially observable networks (single layer obfuscation).
References	[1] https://www.cs.cmu.edu/~badityap/papers/composite-ccr12.pdf [2] http://www.cs.cmu.edu/~deswaran/papers/pakdd17-prone.pdf Sample data: https://www.reddit.com/r/datasets/comments/3mg812/full_reddit_submission_corpus_now_available_2006/

Project III.2	Social Simultaneous Location And Mapping (SocialSLAM) (2 students) [Mentor: Fillipe Condessa]
Overview	How can one efficiently navigate through a network in order to map its general structure? The automatic exploration of social networks through bots and humans is time and resource limited. At the same time, Simultaneous Location And Mapping (SLAM) methods allow autonomous robots to navigate and chart unknown worlds and are pivotal to the autonomous driving advances we observe currently in the streets of Pittsburgh. This project will encompass the development of a framework for efficient navigation in unknown social networks, and resource efficient target (individuals with characteristics of interest) identification in real networks (e.g. Reddit).
Tasks	● Construct static multilayer network from Reddit data (multiple layers); ● Define multiple reward functions modeling quantities of interest (e.g., membership to "dangerous" groups, likelihood of acquiring product, ...); ● Learn inter and intra layer parameters and priors of likelihood of user to hold quantity of interest based on the neighbors' values; ● Explore different heuristics for network navigation with varying costs exploration / exploitation of user's reward function and network structure.
References	[1] https://www.cs.cmu.edu/~badityap/papers/composite-ccr12.pdf [2] http://www.cs.cmu.edu/~deswaran/papers/pakdd17-prone.pdf Sample data: https://www.reddit.com/r/datasets/comments/3mg812/full_reddit_submission_corpus_now_available_2006/

Project III.3	Networks and Dynamics of Political Discussion (2 student) [Mentor: Fillipe Condessa]
Overview	Are the patterns of on-line discussion and communication independent of their contents? How does social interaction differ across the political spectrum? Can political bias be inferred from the patterns of communication from the users? This project allows you to explore the dynamics of social networks of discussion in Reddit, quantify and characterize the patterns of communication in different (political and apolitical) subreddits.
Tasks	● Construct dynamic networks from Reddit data for different sub-reddits representing different politically affiliated groups; ● Identify different network characteristics and patterns of interaction among users affiliated with

	different sub-reddits; ● Analyze the feasibility of political bias inference from the patterns of communication of users and groups of users.
References	[1] https://www.cs.cmu.edu/~badityap/papers/composite-ccr12.pdf [2] http://www.cs.cmu.edu/~deswaran/papers/pakdd17-prone.pdf Sample data: https://www.reddit.com/r/datasets/comments/3mg812/full_reddit_submission_corpus_now_available_2006/

Datasets**:**

- D1: Nightlight satellite data that could be used as a proxy for economic prosperity from NOAA: https://ngdc.noaa.gov/eog/
- D2: Social network (twitter), phone call network data, Weather, etc. for two regions in Italy (Milan and Trentino province): https://dataverse.harvard.edu/dataverse.xhtml?alias=bigdatachallenge
- D3: Italy economic indicator data from surveys: https://www.istat.it/en/archive/193939 or https://www.istat.it/en/products/microdata-files or https://ugeo.urbistat.com/AdminStat/en/it/economia/redditi/trento/22/3#linknote_2
- D4: UK Economic index: http://webarchive.nationalarchives.gov.uk/20120920021427/http://www.communities.gov.uk/publications/communities/englishindices
- D5: Predicting poverty code: https://github.com/nealjean/predicting-poverty/tree/master/data/input
- D6: Living standards measurement study (LSMS), The World Bank: http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html
- D7: The DHS Program (Demographic and Health Surveys): http://dhsprogram.com/
- D8: Change.org API (https://github.com/change/api_docs/blob/master/v1/documentation/index.md).
- D9: http://citibikenyc.com/system-data
- D10: Open data from NYC: http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#bikes or http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#Pedestrians or http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#public or http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#traffic or http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#realtime

Projects:
P I.1: D1, D2, D3, D4
P I.2: D9, D10
P II.1: D1, D5, D6, D7
P II.2: D8

IV. Networked Cyber-Physical Systems

Project IV.1	Large Scale (Biological) Networks Attack (2 students) [TA: Chieh Lo]
Overview	Have you wondered how bacteria can attack the human immune system [1]? By utilizing biological networks such as multiple species of bacteria, we can coordinate their behaviors and attack the human immune system. In this project, you need to first choose your primary target. Next, you need to develop biological ways to effectively attack it.
Tasks	● Choose you attacking target (e.g., T cells or immune system) ● Develop strategies to coordinate your attack on the target (network) ● You should be able to answer: (1) What is the characteristic of the target? (2) What is the best strategy to coordinate the network attack? ● Come up with other interesting questions to answer besides the above. Other tasks include visualizations, coming up with interesting network construction approaches, etc. ● Required: Basic familiarity with MATLAB and Python
References	[1] Antigenic cooperation among intrahost HCV variants organized into a complex network of cross-immunoreactivity. Pavel Skumsa, Leonid Bunimovichb, and Yury Khudyakova. (2014) PNAS. [2] Mathematical analysis of the global dynamics of a model for HIV infection of CD4+ T cells. Liancheng Wang, and Michael Y. Li (2006) Mathematical Biosciences. [3] Cooperation and communication among bacteria: https://en.wikipedia.org/wiki/Quorum_sensing

Project IV.2	From Cellular to Large Scale Computation (2 students) [Mentor: Chieh Lo]
Overview	Are you interested in how to engineer cells to perform certain computational tasks such as adding two integers? Recent advancement in synthetic and systems biology have been directed towards the building of artificial computational devices using engineered biological units as basic building blocks. In this respect, cells can perform basic functions such as AND, OR and NOT logic gates. In this project, you need to engineer cells to implement specific computational tasks at large scale. More specifically, if each cell serves as a basic computational unit (e.g., an AND gate), you need to coordinate cells to accomplish big tasks like pattern formation [3].
Tasks	● Design cells to implement logical functions (e.g., one bacterium can be reprogrammed to implement basic logic gates such as AND, NOT, OR, NAND). ● Implement a more complex function by wiring cells (communication protocols). Show you results evaluated by the performance metrics (e.g., latency, memory, …,etc.) ● Consider the effects of synchronization and lack of synchronization among bacteria communication. What do you observe? Other tasks: Visualization ● Familiarity with MATLAB is required.
References	[1] A. Tamsir, J. Tabor, and C. Voigt. Robust multicellular computing using genetically encoded nor gates and chemical wires". Nature, 2010 [2] M. TerAvest, Z. Li, and L. Angenent. Development of bacteria-based cellular computing circuits for sensing and control in biological systems. Biomolecular Information Processing, 2012 [3] Basu, S. et al. A synthetic multicellular system for programmed pattern formation. Nature. 434, April (2005).

Project IV.3	Distributed DNA Processing: Exploits and Security (2 students) [Mentor: Ryan Kim, Chieh Lo]
Overview	In recent years, big data has influenced genomic sciences, leading to a proliferation of bioinformatic tools. Unfortunately, security practices in these tools have been poor and leave potential for DNA-based exploits. How can these DNA-based exploits attack systems running these tools? If multiple networked machines coordinate to run these bioinformatic tools, how does the network influence the vulnerability of the system? Can we develop safeguards to prevent DNA-based exploits on single-system and distributed system bioinformatic machines? This project will focus on understanding and developing techniques to attack/defend bioinformatic systems using/from DNA-exploits, focusing specifically on how networks of these systems can influence the robustness of the overall operation.

Tasks	● Obtain and introduce a deliberate vulnerability into the bioinformatic tool [1]. ● For the chosen vulnerability, develop attacks based on a single system exploit (application running on a single machine) and a distributed system exploit (application running on multiple machines) ● For the vulnerability, develop detection and prevention/mitigation techniques for attacks of this vulnerability. ● Understand and characterize the robustness and defensibility of these systems while analyzing how the network affects these characteristics
References	[1] https://sourceforge.net/projects/fqzcomp/ [2] http://dnasec.cs.washington.edu/dnasec.pdf

Project IV.4	**How much work can we do at once? Opportunities in distributed systems (2 students) [Mentor: Ryan Kim]**
Overview	Applications, from social modeling to biological simulations, demand more and more data to achieve greater insight into the system they investigate. The simple solution is to distribute these applications across multiple nodes within large computing clusters or data centers. However, for applications operating in an IoT architecture, unreliable or costly network communication makes these types of decisions less obvious. With so many different domains and application requirements a simple "one-shoe-fits-all" mentality does not fit well. How does someone decide the optimal configuration for a distributed application? How much does the communication protocols and network affect the application? How close can we get to the limit proposed by Amdahl's Law? Using a simple Raspberry Pi 3 computing cluster to emulate a distributed IoT edge environment, you will begin to investigate these questions and more.
Tasks	● Find two-three interesting applications that can be easily distributed across a cluster of computing nodes. Choose one application from each of the following categories: Big-Data, a coordinated application (where multiple nodes must cooperate to complete a task), and one of your choice. ● Develop multithreaded C/C++ code for each application. Implement this code across Raspberry Pis using MPI or gRPC ● Discover the relationship between an application's characteristics, an application's performance, and the choices made for distribution and communication across a network of devices. ● Evaluate and optimize the implementations on the basis of application execution time and power consumption for unconstrained and constrained (e.g., power-limited) environments.
References	[1] OpenMPI: https://www.open-mpi.org/ [2] gRPC: http://www.grpc.io/ [3] Raspberry Pi 3: https://www.raspberrypi.org/

V. Create Your Own Project

Project X	Define Your Own Project
Overview	If you have your own idea for a project you'd like to explore in this class, we'd be very happy to hear your proposal! You may bring completely new questions that you would like to investigate; alternatively, you may want to answer the questions similar to the above projects but using other settings and datasets. Either way, please submit a one-page (maximum) proposal by **9/11/2017** carefully and clearly outlining the following:
Tasks	● Project description and motivation. ● Objectives and deliverables (with specific timeline). ● Number of students who will work on the project. ● Availability of datasets (if there is no dataset available, give a viable plan to obtain it) ● Two or more references to related material.
Note	We reserve the right to reject (or significantly modify) your proposals, so please be sure to do some research on the topic beforehand. Of note, your project idea will have a significantly higher chance of being accepted if it follows the same spirit and nature of the ideas embraced by both the course lectures and the other projects already listed. Surprise us with your creativity!
Examples	Below is a project topic suggestion; it is up to you to define the relevant questions to answer or hypothesis to test through network modeling and/or analysis. You still need to submit a "Create your own project" proposal including details of the questions you want to address and your proposed approach of tackling the problem. **Movie/series actor/film studio network:** IMDB, Rotten Tomatoes, and The Movie DB should be good sources to obtain data.