Zeypep Ertem

Research Statement

Executive summary

My research interests are in the **interdisciplinary** side of **operations research** with an emphasis on **challenges in public health and healthcare systems.** Specifically, I am interested in studying operational problems in healthcare systems with policy implications. The recent initiatives in digitization of healthcare records provide an unprecedented measurement capability to how our healthcare systems operate. Leveraging this capability, my research aims to use and extend operations research literature to improve the efficiency of healthcare systems.

US has the largest healthcare budget by annually spending more than 17% of its GDP in the health care system and yet is ranked 37th in overall performance. To an operations research scientist, this suggests an opportunity for improvement. Furthermore, the **availability of patients electronic medical records** when they visit hospitals and **their general health signals through wearable technology devices** such as smartphones and watches enable us to collect unparalleled behavioral data about the experiences of patients in the healthcare system. My research aims to take advantage of these developments to improve the efficiency of healthcare systems.

Developing solutions for healthcare systems face fundamental challenges. (1) Many institutions have started to collect electronic healthcare records but usually, such datasets are owned by proprietary companies. Collaborations with such companies are required to employ such data sources to identify potential improvements in the healthcare system. (2) Solving problems in the intersection of healthcare systems and operational excellence requires medical experts and operational researchers to willingly work with each other. (3) Nudging the decision makers to impact policy changes requires thorough analysis and experimentation with scientific rigor. My research goal is to address these challenges and deliver high-impact solutions to **support decision makers in healthcare systems towards operational excellence.**

My research advances computational and mathematical models that improve our understanding of complex systems in general and healthcare systems in particular. My work is highly interdisciplinary and has made **contributions to both operations research and applied areas of public health.** This is exemplified by my publications in PLOS Computational Biology, Journal of Global Optimization, Social Networks, and Algorithms.

Forecasting outbreaks

One challenge in healthcare resource planning and public safety is timely and accurate forecasts of the severity of infectious disease, especially when the disease level is above the normal expected levels (i.e., disease outbreaks). Many studies use different kinds of data sources to forecast disease outbreaks such as traditional surveillance and internet-based data sources. Yet **systematic evaluation of data sources** to find a powerful subset with the highest forecasting accuracy has been missing. One of my recent published work addresses this challenge through a novel forecasting Bayesian framework with an expert understanding of disease dynamics[1]. This algorithm uses an optimization framework with multi-regression to find a powerful combination of just a few data sources to provide early and accurate outbreak forecasts. Furthermore, this has been **the the first study that also includes electronic health records** as a candidate data source for forecasting thanks to our collaborators in industry. I find that the electronic health records data combined with the traditional surveillance data sources are the most powerful combination in predicting flu outbreaks. Our final framework provides

accurate forecasting (i.e., 15% more accurate than just using the traditional surveillance data) for an army infections disease system called DTRA[1]. This project has been part of a large collaboration between **industry, government, and academia** and I successfully delivered our software solution to the larger team. Furthermore, my forecasting results have been highlighted in several local news outlets[2]. I was not only responsible for developing the algorithm but also for leading the team to integrate our deliverables into the army's digital toolkit.
This project is a stepping stone for using electronic health records in improving operational excellence in healthcare systems. Application of this methodology to other diseases and other countries, as well as using electronic health records to improve the efficiency of healthcare institutions are promising future research directions and highlights **a new research program in the intersection of operations research and healthcare systems** and I am determined to actively move the agenda forward in this program.

Network optimization theory

My research also **contributes to the theory of operations research**. My notable theoretical contributions include two papers. First, a new mathematical model to find cohesive subgroups in networks namely *alpha-cluster*, and a new network clustering algorithm based on *alpha-cluster* [2]. Second, a new theoretical model *Independent Union of Cliques (IUC)* which has close connections to well-known optimization problems like *maximum clique* and *maximum independent set* [3].
Complex systems usually consist of entities that interact with each other through various relationships. For example, people in a society can be friends with each other, researchers in an institution can co-author papers together, or neurons in a brain can work together to perform a certain function. Usually, these types of systems are mathematically modeled using graphs (i.e., networks) where the entities, their attributes and their relationships with other entities can be represented and reasoned about. My research during my Ph.D. studies focused on **developing theory as well as algorithms for novel methods in graph theory and optimization, and their applications to social networks.**
First, I developed a new mathematical model, called *alpha-cluster*, to identify cohesive subgroups which might correspond to a specific group of interest (e.g., social circles in social networks, researchers with similar interests in publication networks, regions performing the same function in brain networks). I employed a widely used graph metric, called *clustering coefficient*, in my model to **capture the small-world property of the many observed real-world networks** (i.e., networks that are highly clustered with small path lengths). I analyzed the theoretical and structural properties of my model along with applying it to several real-life social network instances. Finally, I developed a novel network clustering algorithm based on alpha-cluster that can identify clusters in known networks with the addition of identifying sub-clusters in each cluster.
Second, I developed another new mathematical model, called *maximum independent union of cliques (i.e, max IUC)*, which arises as a special case of alpha-clusters. The model identifies the largest set of nodes that induce independent cliques in a network. This problem has **close connections to well-known optimization problems in the literature such as maximum independent sets, maximum clique, and set covering problems.** For example, any solution to the maximum independent set or to maximum clique problems is a feasible solution for the maximum IUC problem. Also, with a simple modification to the mathematical model, it is straightforward to show that max IUC problem is a special case of the set covering problem. I performed extensive theoretical analysis and developed both exact as well as approximate algorithms to solve this problem. I have employed branch-and-bound algorithms to develop more efficient exact algorithms than off-the-shelf integer programming solvers. Also, I have developed several heuristic-based approaches to find practical approximate solutions.

[1] http://www.dtra.mil
[2] https://www.sciencedaily.com/releases/2018/09/180919154143.htm

Research Agenda

In my future career as a junior faculty member, I plan to advance the area at **the intersection of operations research and healthcare systems.** Specifically, I am interested in achieving operational excellence in healthcare systems using electronic health records and operations research principles. To continue pursuing this agenda, I am excited to investigate three broad problems.

Disease transmissions and outbreaks. Many countries spend significant resources to control and limit the potential effects of disease outbreaks. Timely and accurately predicting the outbreaks enables public health officials to optimize the effects of their prevention strategies. Furthermore, recent refugee movements across countries might pose stress on both the healthcare system and the general public health of the hosting countries. We need to **not only understand the impact to the public health of such population movements but also devise strategies to prevent potential outbreaks.** My recent work in collaboration with industry and government researchers places me in a unique position to address these challenges. I plan to use simulation studies (e.g., SIR models), analysis of real-data involving electronic records with machine learning methods, and mathematical modeling to devise effective strategies for public health.

Hospital operation optimization. Electronic health records also provide data about the operations in healthcare institutions such as hospitals. Such granular measurement capability can be leveraged to identify operational bottlenecks and suggest policies to improve efficiency in the overall system. For example, data-driven decision support models and tools in healthcare systems can deal with the uncertainty in inpatient and outpatient settings resulting in both cost savings for the institution and eventually for the patient as well as increased throughput in the number of patients being served. I want to **use electronic health records data with methods from operations research to have a major impact on hospital operations.**

Scaling discrete optimization to large graphs. My Ph.D. thesis included two major contributions to the theory of clique relaxations by developing two novel graph-relaxations (i.e., *alpha-cluster* and *maximum -IUC*). I have also developed graph mining algorithms such as clustering from these clique relaxations and applied them to problems in social networks and biological networks. However, the computational complexity to solve these mathematical models is high limiting the applicability of these ideas. In my future research, I would like to **develop scalable algorithms to find these clique relaxations and apply such models to large real networks.**

In the long-term, I plan to continue focusing on impactful problems in the **intersection of operations research and healthcare** to improve operational excellence in both healthcare institutions and public health agencies. I have been fortunate enough to collaborate with researchers from diverse academic fields (e.g., operations research, public policy), industry (e.g., Athena Health[3]), and government (e.g., DTRA). I am convinced that by pursuing similar types of collaborations in my future research, I will be able to address challenging research problems in this intersection.

Publications

Zeynep Ertem, Dorrie Raymond, and Lauren Ancel Meyers. Optimal multi-source forecasting of seasonal influenza. *PLOS Computational Biology*, 14(9):1–16, 09 2018.

Zeynep Ertem, Alexander Veremyev, and Sergiy Butenko. Detecting large cohesive subgroups with high clustering coefficients in social networks. *Social Networks*, 46:1 – 10, 2016.

Zeynep Ertem, Eugene Lykhovyd, Yiming Wang, and Sergiy Butenko. The maximum independent union of cliques problem: complexity and exact approaches. *Journal of Global Optimization*, 2018.

[3] https://www.athenahealth.com