A Network-Theoretic Analysis of Hospital Admission, Transfer, and Discharge Data

Maria Cioffi[1], **Naba Mukhtar**[1], **Nathan C. Ryan, PhD**[1], **Joe J. Klobusicky, PhD**[2]
[1]**Bucknell University, Lewisburg, Pennsylvania;** [2]**Rensselaer Polytechnic Institute, Troy, New York**

Abstract

Comprehending complex behavior of flow within a graph is of interest to clinicians and mathematicians alike. In this study we examine admission, discharge, and transfer data of patients within a hospital system, and process the importance of nodes through several graph metrics. One common metric, which measures population densities through a continuous time Markov process, will be compared against centrality measures, a technique more often used in social media studies. Our findings show that centrality measures capture behavior related to the topology of the network that may be missed by Markov processes. This suggests that, for determining the allocation of resources between departments of a hospital, centrality measures in some cases may prove more suitable for interpreting patient flow data. Departmental rankings and suitable instances for the application for each graph metric are provided.

Introduction

A system of communicating agents can often be successfully modeled as a network. One common model studies paths across a network, such as a web session or connecting flights between two cities. Network analysis can consider a large number of paths simultaneously. A natural question, then, is how to create satisfactory descriptors for summarizing flow through a network. For the analysis of flow between departments of a hospital, the subject of this paper, two factors play a key role for how patients transfer across the hospital network. First, there is a time component recovered from departmental transfer and holding times. This is used to describe the population dynamics across the network, which frequently involves the use of Markov models. There is also the overall topology or "shape" of the network. This is more of an issue with the connectedness of the underlying graph, and involves measures of "centrality".

For medicine, the population dynamics of allocation in emergency departments is directly related to quality of care, patient satisfaction, and cost minimization[17]. Dynamic behavior of bacteria and viruses, in settings both inside and outside the hospital, also play crucial roles in understanding the spread of diseases such as malaria or HIV[7, 16, 22, 26]. A widely used and easily interpretable model for hospital populations is the Markov process[4, 9, 10, 13, 15]. In the discrete case, known as a Markov chain, states represent a set of possible patient locations, and (directed) edges are probabilities of transferring from one state to another after one step, which may represent either a fixed time or some other well-defined sequential event (e.g., the rolling a die in a board game). For many processes involving multiple time scales, attempting to define one such step can be too coarse. This is indeed the case when surveying an entire hospital, where differences in the average length of stay are of several orders of magnitude. An alternative approach uses a continuous time Markov process. The associated network for this process assumes that transition rates between nodes are exponentially distributed, with edge weights corresponding to mean transition rates. For a large population occupying a network, densities may be approximated by a differential equation provided through the infinitesimal generator (a type of stochastic derivative) of the Markov process. One advantage to using Markov processes is the wealth of readily applicable results[20]. For instance, in the discrete case, we have simple methods for determining probabilities of transitions after multiple steps. For both the discrete and continuous cases, we can also readily determine limiting distributions, or densities of patients after being allowed to transition for a sufficiently long period.

While Markov processes are helpful in simulations involving population estimates at each node, they do not provide a complete picture of the traffic occurring between nodes. To clarify with a simple example, consider two departments, each containing one patient. In Model A, both patients remain in their initial departments. In Model B, patients transfer between departments at a high frequency. Both models, from a Markovian analysis, will produce the same average density of one patient per department. These densities, however, do not capture interchanges which occur within the system. This consideration is relevant to our model, in which the transfer rate between two departments used in a Markovian analysis can hide important information about diverse traffic. For instance, patients entering operating units in our dataset were quite common, but an operating room holds a single patient, with a typical operating time of a few hours. Visits to neonatal intensive care units, on the other hand, occur less frequently, but may hold multiple

Table 1: Summary statistics for the ADT dataset

Statistic	Value
Number of patients	46,237
Number of rooms	747
Number of departments	66
Total department transfers	305,048
Average number of department transfers	3.94
Average time during encounter	115.42 hours
Average time spent in a department	23.35 hours
Department with most admissions	Emergency medicine (42.19%)
Proportion of patients with at least one transfer	83.93%
Department with longest average stay	Geriatric psychology (11.84 days)
Department with shortest average stay	Preoperative waiting rooms (15.85 minutes)

infants, with holding times spanning up to several weeks. The variables in patient traffic have direct applications to epidemiology, which is highly interested in how quickly an infected agent travels across a population.

To account for nodes which may have fewer patients at a given moment, but see more traffic by virtue of how connected they are to other nodes in the network, we employ a variety of centrality metrics for networks. As these metrics consider how connected a node is to other nodes, it is fair to describe them as measurements of certain aspects of the shape of the network. These metrics were largely used in mathematical sociology in the 1970s[2,11], but have recently enjoyed a rebirth in computer science with the advent of search engines and social media[12,23]. The purpose for developing different centrality methods for a given network stems from the many ways to consider a node's importance. We may, in one case, be interested simply in the number of connections of a particular node. In another case, we might instead wish to find which nodes serve as "hubs", or are intermediates in the shortest paths between a large number of nodes. In this paper, we have decided on degree, closeness, betweenness, and eigenvector centrality, which together provide a wide variety of methods for ranking nodes.

In *Data and Methods*, we give a brief overview of each centrality metric, and how each such metric measures a node's importance. We also review continuous time Markov processes, their connection to differential equations for describing population flows, and how to simulate solutions. Each method produces a unique ranking of nodes, which are presented in the *Results* section. The implications of the difference between the various metrics are given in the *Discussion* section. Finally, we summarize our findings in the *Conclusion* section.

Data and Methods

Our dataset consists of time-stamped variables relating to ADT (admissions, discharges, and transfers), consisting of 46,247 patients admitted between December 2013 and March 2016 who permitted the use of their data for research purposes. All patients were seen at one of three hospitals (Geisinger Medical Center, Geisinger Wyoming Valley, and Geisinger Shamokin Area Community Hospital) operated by Geisinger Health System, located in northeastern and central Pennsylvania. Patients transferred through a total of 747 rooms, each of which is assigned to one of 66 departments, such as telemetry units, operating rooms, nurseries, etc. A summary of descriptive statistics is given in Table 1. For the entirety of this paper, we consider statistics for all three hospitals taken together, including transfers between hospitals. As we will see in the methods section, however, after building an appropriate network to describes patient flows, it is easy to restrict analysis to a specific hospital by simply running the necessary algorithms on a subgraph.

Basic Network Theory Terminology

A *network* is a collection of nodes and links between them. A hospital (or more generally a healthcare system) can naturally be modeled as a network. For example, the nodes could be the departments in the hospital (or the rooms, or the doctors) and we could place a link between node A and node B if at some point a patient moved from room A to room B or vice versa. A network is said to be *directed* if links go from one node to the other but not necessarily the other way. In the same example as above, we would add a link from node A to node B if at some point at least one

Figure 1: A visual representation of the hospital department network. Each node corresponds to a hospital department, and an edge between nodes signifies a patient transfer between departments.
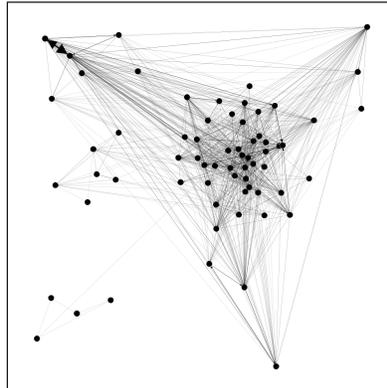
patient moved from room A to room B. A network is said to be *weighted* if a real number is attached to each link. Still with the same example, we could place a link between node A and B with weight 11 to represent that 11 patients moved from room A to room B at some point. See Figure 1 for an example of a weighted and directed network (all network visualizations in this paper are done using Gephi[1]). This particular weighted directed network is the network that we use to represent the Geisinger Medical Center data we use in our study. Suppose a weighted directed network $S = (N, L)$ has nodes $N = n_1, \ldots, n_j$ and links $L = \ell_1, \ldots, \ell_k$. Then, since it is directed, the links are *ordered* pairs (ℓ_r, ℓ_s) corresponding to an link from node n_r to node n_s. Since S is weighted, each link ℓ_t has associated to it a real number called its *weight* that we denote by $\mathrm{wt}(\ell_t)$. We denote the set of nodes of S by N and its set of links by L. We use $\#N$ to denote the number of nodes.

For each node in a network, we say that its *degree* is the number of links going into or out of the node. Analogously, a node's *in-degree* is the number of links going into the node and its *out-degree* is the number of links leaving the node. In a weighted graph, a node's *strength* is the sum of the weights of the links either leaving or entering the node and the analogous ideas of *in-strength* and *out-strength* can be defined, respectively, as the sum of the weights of the links entering and leaving the node. For the network S described above, then, we have the following definitions. Let n be a node of S, then

$$\mathrm{ideg}(n) = \sum_{(m,n)\in L} 1, \qquad \mathrm{odeg}(n) = \sum_{(n,m)\in L} 1, \qquad \mathrm{deg}(n) = \mathrm{ideg}(n) + \mathrm{odeg}(n);$$

$$\mathrm{istr}(n) = \sum_{\ell=(m,n)\in L} \mathrm{wt}(\ell), \qquad \mathrm{ostr}(n) = \sum_{\ell=(n,m)\in L} \mathrm{wt}(\ell), \qquad \mathrm{str}(n) = \mathrm{istr}(n) + \mathrm{ostr}(n).$$

For two nodes in a network, a *path* from one to the other is a sequence of links of the network that start at the first node and end at the second node. The *shortest path* between two nodes on a weighted network is the path with the smallest sum of weights of its edges.

Markov Processes

To understand the evolution of the total number of patients, we use a continuous time Markov process. Here, nodes (often called states in the Markovian context) are given as rooms and departments. Links between states are weighted and directed, with edge weights a_{ij} corresponding to constant rates from state i to state j. This can be converted to a probabilistic interpretation as follows. When considering a closed system, in which all patients are confined to a given collection of states, the evolution of patient density $p(t) = (p_1(t), p_2(t), \ldots, p_n(t))$, where $p_i(t)$ denotes the fraction of patients in room i at time t, is

$$\frac{dp_i}{dt} = \sum_{k|k\to i} a_{ki} p_k - \sum_{k|i\to k} a_{ik} p_i.$$

(A similar model has been used[14, 27] in the field of self assembly in nanoscience). This can be written compactly as $\frac{dp}{dt} = Ap$ with matrix entries a_{jk} for $j \neq k$, and $a_{jj} = -\sum_{k|i \to k} a_{ik} p_i$.

For the problem of flow through a hospital, however, we must also deal with both patients who are admitted or released. Patients who leave a hospital may be represented by the inclusion of an absorbing, or "cemetery", state. Handling admissions is more complex, and involves a patient "influx vector" $c = (c_1(t), \ldots, c_n(t), 0)$ which gives rates of admission to various rooms, with the final 0 entry corresponding to the cemetery state. For the augmented matrix $\tilde{A}$ containing an extra row and column for the absorbing state, the evolution equation for total number becomes

$$\frac{dp}{dt} = \tilde{A}p + c(t)$$

To compute coefficients, we considered rates between two states in patients per minute. Precisely, for a proportion $q(X, Y)$ of all patients in state X which transfer to state Y, having times $t_1, \ldots, t_n$ spent in state X, the transition rate is given as the average

$$a(X, Y) = q(X, Y) \sum_{i=1}^{n} \frac{1}{t_i}.$$

Along with using the empirical rates just decribed, we will also consider a uniform simplification where patients, on average, spend equal time in each state. From this assumption, transition rates simplify to $a(X, Y) = nrq(X, Y)$, where r is the mean time spent, where we average over all departments. Note that this cruder ranking method should be seen as comparable to the other centrality metrics, which are more concerned with connectivity, rather than actual transfer times.

For simulation of the differential equation (), we may employ Duhamel's formula, which provides the explicit solution

$$p(t) = e^{\tilde{A}t} p(0) + \int_0^t e^{\tilde{A}(t-s)} c(s) ds, \quad t > 0,$$
$$p(0) = p_0,$$

where e^M denotes the matrix exponential[5] of a matrix M. For our purposes, we consider initial conditions of an empty hospital ($p_0 \equiv 0$), although this methodology allows for running simulations on multiple scenarios, such as a sudden admission spike in the emergency department. With a continuous influx of patients, we expect, and in fact obtain, convergence of populations for each department, since total transition rates to the cemetery state approximately match total influx into the hospital, and all patients are eventually discharged.

Centrality Measures

From the perspective a network's connectedness, it is natural to ask which nodes are "most connected" or "most central". Interpretations of centrality vary. Freeman[11] identified three such interpretations by defining notions of degree, closeness, and betweenness centrality. Bonacich[2] identified a fourth centrality measure that we also consider, the so-called eigenvector centrality.

In a weighted directed network, the *in-strength (respectively, out-strength) centrality* of a node takes into the strength of the node and not solely its degree. We feel justified in only considering this centrality measure because we are trying to understand patient flow in a hospital system and the number of patients flowing through a node. We do mention that Opsahl-Agneessens-Skvoretz[21] introduced a tuning parameter that allows one to weight the strength and degree differently, depending on what one wants to measure. Formal definitions of these centrality measures are as follows. Let n be a node in a network $S = (N, L)$ with

$$\text{in-strength centrality of } n = \text{ISC}(n)$$
$$= \text{istr}(n),$$
$$\text{out-strength centrality of } n = \text{OSC}(n)$$
$$= \text{ostr}(n).$$

In a weighted directed network, the *closeness centrality* of a node is the reciprocal of the node's farness, or the sum of the node's distance to every other node. Newman[18] provided a suitable translation of "distance" when the graph is weighted: attach a *cost* (often the reciprocal of the weight) to the node in question and then that node's distance to another node is the smallest total cost among all paths from the node in question to the destination-node. Again, a variant has been proposed[21] that also takes into account the number of links in the least costly paths. These algorithms used Dijkstra's famous algorithm[8] for finding shortest paths on directed networks as their starting points. Now we provide a formal definition: Let n be a node and let $c(m, n)$ be the cost of the link from m to n, that is, in our context, the reciprocal number of patients going from node m to n. A smaller cost means that the two nodes are more closely connected. Then the closeness centrality of n is defined as

$$\mathrm{CC}(n) = \frac{1}{\sum_{n \neq m \in E} c(m, n)}.$$

In a weighted directed network, the *betweenness centrality* of a node is a measure of how much of a role the node has in the flow among and between the other nodes in the network. This algorithm can be modified[21] by the introduction of a tuning parameter to allow one to take the number of links involved in the paths and not just the costs. Let n be a node in $S = (N, L)$ and set $\sigma(a, b)$ to be the number of least costly paths from node a to node b and let $\sigma(a, b \mid n)$ be the number of least costly paths from node a to b that pass through node n. Then the betweenness centrality is given as

$$\mathrm{BC}(n) = \sum_{a, b \in N} \frac{\sigma(a, b \mid n)}{\sigma(a, b)}.$$

In a weighted directed network, the *in-eigenvector centrality* of a node is a measure of its importance determined by the importance of the nodes that link to it and those that link to them and so on. Let $A = (a_{ij})_{1 \leq i, j \leq \#N}$ be the adjacency matrix of the network $S = (N, L)$. That is, the entry in the ith row and j^{th} column is $\mathrm{wt}(n_i, n_j)$, the weight of the link from i^{th} node to the jth node or zero if there is no such link. If the network satisfies certain technical conditions (irreducibility), then, by the Perron–Frobenius Theorem, A has $\#N$ distinct eigenvalues, with a unique largest one λ_1 (see Thm 1.3.5 in Kitchens[28]). The left eigenvector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{\#N} \end{pmatrix}$ corresponding to λ_1, the unique solution of the matrix equation

$$\mathbf{x}A = \lambda_1 \mathbf{x},$$

encodes the in-eigenvector centrality. That is

$$\text{in-eigenvector centrality of } n = \mathrm{IEC}(n) = x_i,$$

where n is the ith node in S and x_i is the ith component of $\mathbf{x}$. To find the out-eigenvector centrality, one starts with the transpose of A and finds the right eigenvector of A^t; for a node n, we denote that measure by $\mathrm{OEC}(n)$.

Our centrality results are computed using the implementations of these centrality measures in NetworkX[24] and custom written code available online[6].

Results

We summarize the results of our calculations in a series of tables and figures. To determine a ranking of departments, we take average populations of departments estimated from the Markovian model over a period of 1000 days. All departments approach a limiting distribution for their populations. Convergence times differ for each department, but all settle by day 30. The convergence of popular departments for each of our Markov models are given in Figures 2–3.

In Table 2 we identify those nodes that are most important according to our centrality metrics described above.

Discussion

Above we have alluded to some of the interpretations for the various metrics. We now consider consider each of these interpretations and identify which metrics are better suited for answering appropriate questions from the perspective of hospital flows.

Figure 2: Left: Prevalent departments under a continuous time Markov chain, modeled as patients in a department under a constant influx of admitted patients. Note that for this graph we have used log scaling for the dependent variable. Right: Prevalent departments under a continuous time Markov chain modeled as the number of patients in a department under a constant influx of admitted patients, assuming constant transition rates between connected nodes.
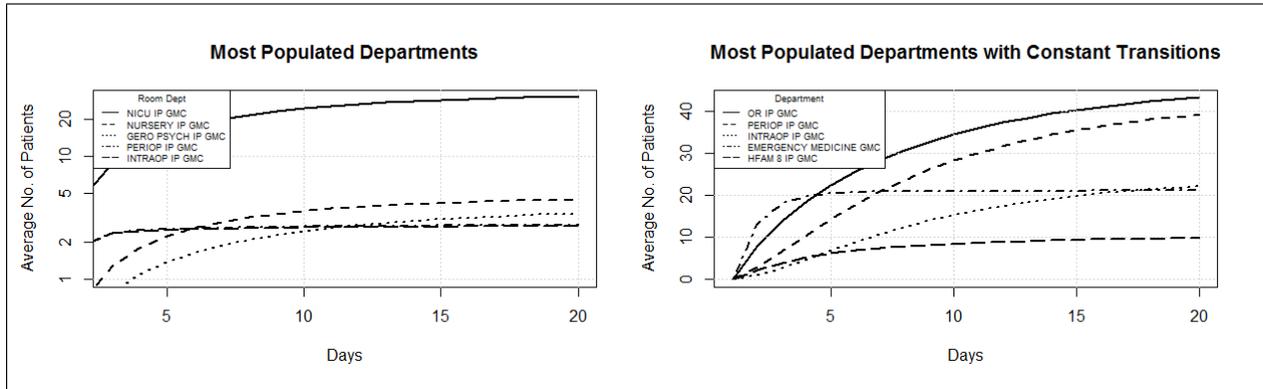

Figure 3: Prevalent individual rooms under a continuous time Markov chain, modeled as patients in a department under a constant influx of admitted patients.
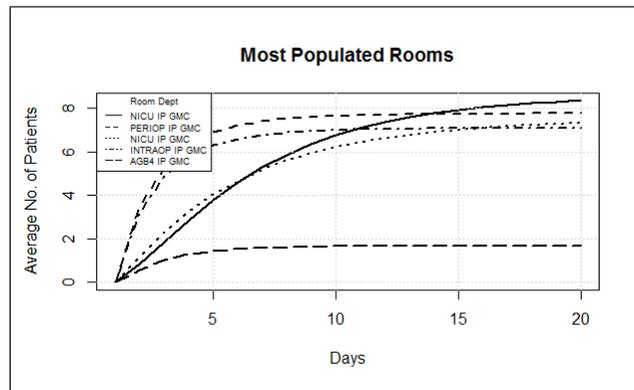

Table 2: The top four departments by various centrality measures. We recall that ISC is in-strength centrality, OSC is out-strength-centrality, CC is closeness centrality, BC is betweenness centrality, IEC is in-eigenvector centrality and OEC is out-eigenvector centrality.

	ISC[a]	OSC	CC
1	PERIOP IP [b]	OR IP	EMERGENCY MEDICINE GSACH[c]
2	OR IP	PERIOP IP	EMERGENCY MEDICINE
3	INTRAOP IP	EMERGENCY MEDICINE	OR IP
4	HFAM[d] 8 IP	INTRAOP IP	PERIOP IP

	BC	IEC	OEC
1	OR IP	PERIOP IP	PERIOP IP
2	HFAM 8 IP	OR IP	OR IP
3	MED SURG - WEST WING IP GSACH	INTRAOP IP	INTRAOP IP
4	BP7[e] IP	HFAM 6 IP	IN OUT SURGERY 2 IP

We mention that three departments consistently appeared as the most important nodes for most metrics. Those departments were the Operating department, the Perioperating department and the Intraoperating department. In this section we will also justify why the list of other departments that appear in our results is not surprising.

We also point out that these results are for regional hospitals in Geisinger Health Systems, but in general results may be hospital dependent. For instance a community hospital or a teaching hospital might perform different kinds of procedures and have different expectations of how long a procedure might take.

Markov models The continuous time Markov process makes predictions based on population densities. This takes into account a patient's average length of stay and also instantaneous populations for a room. This population based ranking would heavily weight departments that can hold a large numbers of people at a particular time. Therefore, the Neonatal ICU, Nursery and Geriatric Psychology departments are those that are most likely to be populated with the largest number of people at a given time. Trivially, the most important room is the absorbing state, as every patient is eventually discharged. Our results, therefore, scale populations assuming that they have not exited the hospital.

The most classical applications of Markov processes would address issues of patient overflow. The closed form steady state analysis of this paper provides one such method for ranking populations. However, a more direct approach would be through applying "worst-case scenario" type initial conditions to the predetermined transition matrix, such as an influx of patients into an emergency department following a catastrophe.

By imposing constant transition rates to the continuous time Markov model with constant transition rates makes predictions of where the traffic through the network might bottleneck. By considering all time-steps to be the same size, accurate predictions of a node's population are lost, but nodes that score highly in this metric are considered to be, on average, those which do not clear people out as quickly as they take them in. That the top three nodes according to this metric are the ones consistently identified, as we will see below, by the other "topological" metrics is not surprising.

In- and out-strength centrality In-strength centrality measures the number of patients who have flowed into the node considered over a dataset's time range. The top three rooms for in-strength centrality are the Perioperative department, the Operating department and the Intraoperative department. These are not surprising because these departments serve as a hub, accepting multiple patients from a variety of departments. In general, in strength centrality measure a department's "receptivity" to other departments taken as a whole.

Similarly, out-strength centrality measures the number of patients flowing out of the department. A node with a few heavy-weighted links out of it or one with many light-weighted links out of it would have a large out-strength centrality score. Again, three of the top four are the Operating department and the Intraoperative department. The other one in the top four is Emergency Medicine, which is reasonable, as an Emergency Medicine department often admits patients to other departments.

Closeness centrality Department A and Department B (rooms, beds, etc.) are close, in the sense of closeness centrality, if many patients transfer between departments A to B. A department with a large closeness centrality score is then one which is close many departments. Our results show that two Emergency departments are highly ranked. In general, closeness centrality may help find those departments in which patients will eventually travel to many other departments. This differs slightly from in and out strength centrality, in that closeness centrality is not local in the sense that the ease of traveling through a graph considers paths of length larger than 1.

Betweenness centrality A department A is between two departments X and Y if, among all ways to get from department X to department Y, a large number of paths go through A. So, A would have a large betweenness centrality score if A is highly between all pairs of nodes. That is, a large number of patients flow through node A and A is the most efficient (in terms of not diverting population away from their main path) of all nodes. The rooms identified in the results play exactly that role in the hospital.

There is a possible connection with a betweenness score and the spread of infections, such as C Diff., in a hospital. A room which has not been properly disinfected may be more damaging if it is a common room that many patients pass

through. This may be contrasted with a room that at any time may contain many patients, but is isolated from other departments.

In- and out-eigenvector centrality   A department with high in-eigenvector centrality is one that linked to many other nodes which are also highly connected. The three departments that score most highly in this metric, namely the Operating, Perioperating, Intraoperating departments, are ones with have a lot of long and highly weighted paths that terminate there. It might seem, then, that the nodes with largest in-eigenvector centrality are the same as those that score highest using the Markov chain metric. The difference, though, is that in the Markov model a patient can stay in a department over two time steps, making it a good measure of how where there might be bottlenecks. In the in-eigenvector centrality, we are measuring which departments receive a lot of patients from departments with many patients, which themselves receive many patients, and so forth.

A department with high out-vector centrality is one that links to other departments with high out-vector centrality and which are themselves linked to departments with high out-vector centrality, etc. Departments with large in-eigenvector centrality, like those with large Markov Chain steady states, represent bottlenecks. Whereas important departments from the Markov Chain encode some coarse information about time (e.g., a patient might stay in a node for several time steps), the in-eigenvector centrality does not. The departments with large in-eigenvector centrality, then, identify departments that are bottlenecks because of the number of steps involved, ignoring how long the steps might take. Identifying departments that score high in this metric would allow administrators another opportunity to disperse bottlenecks.

Conclusion

In this work, we have applied several graph theoretic techniques for determining the relative importance of hospital departments. Each technique carries a unique interpretation, which we found to align with intuition regarding traffic in hospitals. For instance, emergency departments are highly prevalent in closeness centrality, which roughly measures ease of access to other departments, and nurseries are highly ranked in a Markovian analysis, due to average prolonged waiting times which increase average population. We find, from the variety of prevalent nodes arrived at through different graphical centrality measures, that a full description of a patient flow in a hospital may require multiple techniques. This is especially true when considering more general networks, in which a large number of scales may occur, both in time and state (e.g. rooms versus departments as nodes in a graph).

While a common theme in operations research involves Markovian analysis to answer questions regarding population, we hope that a deeper investigation of alternative centrality measures can reveal information pertaining to deeper questions of queueing theory such as bottlenecking or rates or communicability between departments, either as a result of protocol or practitioner error. We also hope in future work to address how centrality measures may relate to the optimized care regarding histories of an individual patient post-surgery. Specifically, we will look to find associations between hospital acquired infections and certain departmental pathways which are weighted by importance obtained through the methods discussed here.

References

1. Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

2. Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.

3. Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

4. James R Broyles, Jeffery K Cochran, and Douglas C Montgomery. A statistical markov chain approximation of transient hospital inpatient inventory. *European Journal of Operational Research*, 207(3):1645–1657, 2010.

5. Carmen Chicone. Ordinary differential equations with applications. 2006.

6. Maria Cioffi, Joseph Klobusicky, Naba Mukhtar, and Nathan C. Ryan. Code for "network-theoretic analysis of a hospital to study patient populations, 2016.

7. John M Coffin. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, 267(5197):483, 1995.

8. Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.

9. MJ Faddy and SI McClean. Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry*, 15(4):311–317, 1999.

10. MJ Faddy and SI McClean. Markov chain modelling for geriatric patient care. *Methods Archive*, 44(3):369–373, 2005.

11. Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

12. Rumi Ghosh and Kristina Lerman. Predicting influential users in online social networks. *arXiv preprint arXiv:1005.4882*, 2010.

13. Edward PC Kao. Modeling the movement of coronary patients within a hospital by semi-markov processes. *Operations Research*, 22(4):683–699, 1974.

14. Ryan Kaplan, Joseph Klobušickỳ, Shivendra Pandey, David H Gracias, and Govind Menon. Building polyhedra by self-assembly: theory and experiment. *Artificial life*, 2014.

15. Peter Kolesar. A markovian model for hospital admission scheduling. *Management Science*, 16(6):B–384, 1970.

16. K Magori and JM Drake. The population dynamics of vector-borne diseases. *Nature Education Knowledge*, 4(4):14, 2013.

17. Megan McHugh, Kevin VanDyke, Mark McClelland, and Dina Moss. Improving patient flow and reducing emergency department crowding: a guide for hospitals. 2012.

18. Mark EJ Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132, 2001.

19. Hiro-Sato Niwa. School size statistics of fish. *Journal of theoretical biology*, 195(3):351–361, 1998.

20. James R Norris. *Markov chains*. Number 2. Cambridge university press, 1998.

21. Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social networks*, 32(3):245–251, 2010.

22. Andrew Pearson. Population dynamics and disease. *Proceedings of the Royal Society of Medicine*, 59(1):55, 1966.

23. Hassan Sayyadi, Matthew Hurst, and Alexey Maykov. Event detection and tracking in social streams. In *Icwsm*, 2009.

24. Daniel A Schult and P Swart. Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, volume 2008, pages 11–16, 2008.

25. RA Sunyaev and Ya B Zeldovich. Formation of clusters of galaxies; protocluster fragmentation and intergalactic gas heating. *Astronomy and Astrophysics*, 20:189, 1972.

26. Glenn F Webb, Erika MC D'Agata, Pierre Magal, and Shigui Ruan. A model of antibiotic-resistant bacterial epidemics in hospitals. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13343–13348, 2005.

27. Adam Zlotnick. To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *Journal of molecular biology*, 241(1):59–67, 1994.

28. Bruce P. Kitchens. Symbolic dynamics: one-sided, two-sided and countable state Markov shifts. 2012.