# Spread of pathogens in the patient transfer network of US hospitals

Juan Fernández-Gracia[1,2], Jukka-Pekka Onnela[3], Michael L. Barnett[3], Víctor M. Eguíluz[2], and Nicholas A. Christakis[4]

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health
[2] Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB)
[3] Department of Biostatistics, Harvard T.H. Chan School of Public Health
[4] Yale Institute for Network Science, Yale University

**Abstract.** Antibiotic-resistant organisms, an increasing source of morbidity and mortality, have a natural reservoir in hospitals, and recent estimates suggest that almost 2 million people develop hospital-acquired infections each year in the US alone. We investigate the temporal network of transfers of Medicare patients across US hospitals over a 2-year period to learn about the possible role of hospital-to-hospital transfers of patients in the spread of infections. We analyze temporal, geographical, and topological properties of the transfer network and show that this network may serve as a substrate for the spread of infections. Finally, we study different strategies for the early detection of incipient epidemics on the temporal transfer network as a function of activation time of a subset of sensor hospitals. We find that using approximately 2% of hospitals as sensors, chosen based on their network in-degree, with an activation time of 7 days results in optimal performance for this early warning system, enabling the early detection of 80% of the *C. difficile*. cases with the hospitals in the sensor set activated for only a fraction of 40% of the time.

Every year in the US alone, there are 1.7 million nosocomial infections and 99,000 associated deaths, imposing substantial clinical and financial costs to the US health care system [1–3]. The vast majority of these are due to antibiotic-resistant bacteria [4], which have a natural reservoir in hospitals, presenting a potentially lethal threat to already-sick patients. The annual cost of antibiotic-resistant infections in the US has been estimated to range from $21 billion to $34 billion [5–7]. A 2013 CDC (Centers for Disease Control and Prevention) report on antibiotic-resistant bacteria identified the lack of infrastructure to detect and respond to emerging resistant infections as a pressing gap.

Antibiotic-resistant organisms have a natural reservoir in hospitals. In our study, over a two-year period, there were nearly one million transfer events across US hospitals of Medicare patients alone. Given this large number of transfers, the network of patient transfers could plausibly act as a conduit for antibiotic-resistant bacteria from hospital to hospital. There are, however, only a few existing studies that have investigated the possible role of hospital-to-hospital transfers of patients for the spread of infections. Some studies have focused on the

structure of the nationwide transfer network associated with critical care [8–11], while others have had a more restricted scope, limited to smaller geographical units, such as counties [12, 14].

Local containment of antibiotic-resistant bacteria at the level of individual hospitals is a difficult but manageable task given that interactions between hospital wards are relatively structured and confined spatially [15, 16]. But controlling a larger epidemic of antibiotic-resistant bacteria or responding to new mass outbreaks is much more challenging. This is in part related to the complex pattern of patient movements between hospitals, which gives rise to a broad, distributed network. To better understand the role of patient transfers for the spread of infections, we pursue three interconnected aims. First, we investigate the structure of the hospital-to-hospital patient transfer network in the US; second, we correlate the incidence of nosocomial infections on a national scale with properties of this network; and third, we develop a scalable method for the efficient early detection of the spread of nosocomial infections.

# 1 MATERIALS AND METHODS

## 1.1 Study data

We study hospital-to-hospital transfers of the entire population of US Medicare patients over a two-year period. Medicare provides almost universal coverage to all Americans aged 65 and older, about 15% of the US population [17]; and about 37% of all hospital admissions in 2003 were for Medicare patients [18]. We used a 100% sample of the Medicare Provider Analysis and Review (MedPAR) files for calendar years 2006 and 2007. The MedPAR files contain diagnosis, procedure, and billing information on all inpatient and skilled nursing facility (SNF) stays. Our study cohort consisted of Medicare patients aged 65 or older with a hospital stay at an acute medical or surgical hospital with an active record in the American Hospital Association (AHA) 2005 database [19]. Before applying these exclusion criteria, we identified 26.4 million stays of 12.5 million patients in 6,278 different hospitals. After the exclusions, our final cohort consisted of 21.0 million inpatient stays of 10.4 million patients in 5,667 different hospitals.

## 1.2 Hospital-to-hospital transfers

According to our definition, a hospital-to-hospital transfer occurs whenever a patient is discharged from one hospital and admitted to another hospital on the same calendar day. Note that a minority of transfers as defined here may not correspond to actual formal transfers of patients. For example, a patient could be discharged from hospital A and then be re-admitted to hospital B on the same day for a reason that is unrelated to her stay at hospital A. From an epidemiological point of view, however, these are essentially equivalent to formal patient transfers. Using this definition of transfer, we identified 936,101 transfer events taking place between 76,003 pairs of hospitals.

### 1.3 Constructing the transfer network

We consider a network representation of the patient transfers across hospitals. Hospitals are represented as nodes and a transfer of a total of $x$ patients on day $d$ from hospital $i$ to hospital $j$ is represented as a directed edge from node $i$ to node $j$ with weight $x$ on day $d$. The longitudinal sequence of patient transfers forms a directed, weighted, temporal network. We consider a static representation of the network that retains no temporal information of patient transfers by aggregating the data for the two-year period, where the weight of the edge from node $i$ to node $j$ is the mean daily number of patient transfers through that edge, *i.e.*, the total number of transfers from hospital $i$ to hospital $j$ during the study period divided by the number of days in the period (730 days).

### 1.4 C. difficile incidence on the transfer network

The MedPAR files contain diagnosis codes for each patient. We investigated the incidence of *Clostridium difficile* (*C. difficile*) infections and its correlation with properties of the transfer network. *C. difficile* is an anaerobic, gram-positive, spore-forming bacteria that occurs frequently in health care settings, found in over 20% of patients hospitalized for more than one week. The disease is spread by ingestion of *C. difficile* spores, which are very hardy and can persist on environmental surfaces for months without proper hygiene [20]. *C. difficile* associated infections kill an estimated 14,000 people a year in the US as a result of institutional infections [21]. We ascertained incident cases of *C. difficile* infection by identifying any hospital admissions with ICD-9 diagnostic code 008.45. The sensitivity and specificity of using ICD-9 codes to identify *C. difficile* infections have been reported by multiple groups to be adequate for identifying overall *C. difficile* burden for epidemiological purposes [22–24].

### 1.5 Sensor placement on the hospital network

To set up a real-time surveillance system for infections, such as a new strain of antibiotic-resistant *C. difficile* . It is unlikely that exhaustive data would be available for all hospitals all the time, and this limitation calls for a parsimonious approach where only a subset of hospitals needs to be monitored at any given time. We call these monitored hospitals "network sensors" in the sense that they could be used to sense incipient epidemics. We consider three different prescriptions for sensor placement: (1) choose sensor hospitals in proportion to their in-degree rank in the static network; (2) choose sensor hospitals in proportion to their out-degree rank in the static network; and (3) choose sensor hospitals uniformly at random from the set of all hospitals. In our simulations, we assume that a monitored hospital is able to detect every infected patient who is present either in the hospital itself or in any of its network neighbors to which it is connected via patient transfers. To learn about the potential of the hospital sensor framework to detect epidemics, we investigate its best-case performance by determining the optimal sensor set for the observed data.

### 1.6 Determining the optimal sensor set

We define the relative efficacy of the sensor $E_N$ set as $E_N = D_N/ND_1 - (M - D_N)/M$ where $N$ is the number of sensors in the sensor set, $D_N$ the number of infected patients detected by a sensor set of $N$ sensors, and $M$ is the total number of *C. difficile* cases in the network. While adding sensors to the system always improves its overall performance, any sensor set exhibits diminishing marginal returns in the sense that the per-sensor increment in performance declines with each added sensor. The first term in the definition corresponds to the number of detected cases normalized by the number of cases that would be detected if all sensors were as efficacious as the first sensor in the sensor set. The second term is a penalty term that corresponds to the fraction of undetected cases. High relative efficacy is therefore a combination of selecting a set of sensors that are as close as possible to the efficaciousness of the first sensor in the set and having these sensors miss as small a proportion of cases as possible. Note that the two terms in the definition of the relative efficacy could be assigned different weights; however, here, we opted for the simplest approach and only ensured that the two contributions are measured on the same scale.

### 1.7 Implementation of network sensors

The sensor hospitals can be either passive or active. When a sensor is passive, it can only detect infections in the hospital itself. Whenever an infection is detected, the sensor either transitions from the passive state to the active state for a period of $T$ days or, if already in the active state, remains in that state for another $T$ days. In addition to the efficacy of the sensor sets, for both implementations, we keep track of the fraction of *C. difficile* cases that are detected in order to assess the performance of the sensor system.

We monitor the admission times of *C. difficile* patients at each hospital, and whenever such a patient is admitted, we incorporate the hospital in the sensor set for $T$ days, the activation time, following the admission. Once added to the sensor set, the hospital can detect the *C. difficile* cases present in the hospital itself and its network neighbors for a total of $T$ days. The efficacy of the sensor system therefore depends on the value of $T$, and we compute the efficacy of the sensors for $T$ from 0 to 100 days (shown from 0 to 30 days in Fig. 2 left). For each combination of parameter values, the number of sensors and the activation time, and for each strategy of prescribing sensors, we perform 1,000 independent realizations of the sensor selection process. We also track the average time each sensor stays in the active state. An optimal sensor set is one that has maximal efficacy for activation time $T$, minimizes the average time the sensors stay active, and maximizes the fraction of detected cases.

## 2 RESULTS

### 2.1 Properties of the transfer network

The topology of the network and the geography of patient transfers are closely related, with 90% of transfers between hospitals less than 200km apart (Fig.1 left). On average, over the 2-year period, a hospital sent patients to 13.55 ± 0.15 (SE) hospitals and received patients from 13.55 ± 0.25 hospitals (note that the two means necessarily coincide in a directed network). The average number of patients transferred per edge in the 2-year period was 12.3 ± 0.63 (SE). Although the degree distributions (in-degree and out-degree) have fat tails (more so the in-degree), comparisons of the average clustering coefficient and the average shortest path length to randomized versions of the network show that the network closely resembles a spatial network. In particular, it is much more clustered than a random network and has a high average shortest path length. Finally, the network shows no significant assortativity by degree.

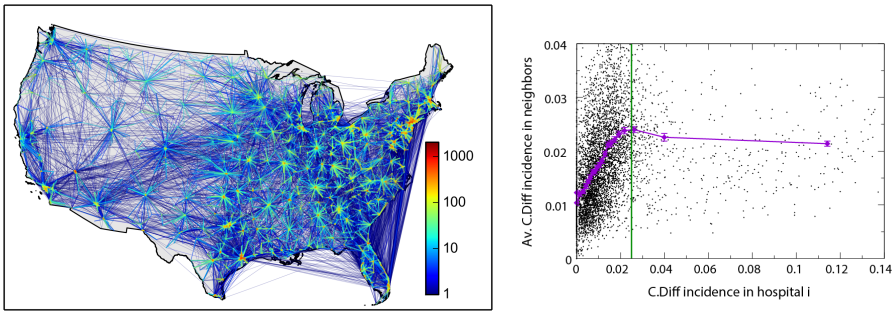


**Fig. 1. Left: Hospital transfer network of US Medicare patients.** The network consists of hospitals connected by daily transfers of patients, aggregated over the two-year period. Edge color encodes the number of patients transferred through each connection. **Right: Correlation between *C. difficile* incidence and transfer network structure.** The x-axis represents the temporal *C. difficile* incidence at the focal hospital over time and the y-axis is the mean *C. difficile* incidence in its network neighborhood (the mean taken first over time and then over all network neighbors). We exclude hospitals with fewer than 100 patients from subsequent correlation analyses, leading to exclusion of 7.5% (428) of all hospitals. The Pearson correlation coefficients are 0.47 and -0.01 for the low and high incidence regimes, respectively, which are separated by the vertical line.

### 2.2 Spread of C. Diff. infections

Over the two-year period, there were a total of 313,214 *C. difficile* infections in the 5,677 hospitals included in the study. The mean *C. difficile* incidence for each

hospital and the mean *C. difficile* incidence for its network neighbors show two distinct regimes, one for low *C. difficile* incidence and another for high incidence (Fig.1 right). The incidence of the pathogen in a given hospital is correlated to the incidence of the pathogen in its network neighborhood if the incidence at the focal hospital is relatively low; this correlation appears to vanish for hospitals displaying higher *C. difficile* incidence. One explanation for this phenomenon is that, if there were only very few cases of *C. difficile* in the low incidence regime, the transfers of infected patients might go undetected, therefore inducing correlations among pathogen incidences across the network. Conversely, if pathogen incidence were high and local, such that hospital outbreaks are detected, patient transfers might be restructured to curb the further spread of the infection. We determine the boundary between the two regimes based on the strength of correlation in pathogen incidence and assign the value for the crossover between the two regimes (shown as the vertical line in Fig. 1 right). For *C. difficile* incidence below this threshold, the Pearson correlation coefficient $R \approx 0.47$ (95% CI: 0.44, 0.49) whereas above the threshold $R \approx -0.01$ (95% CI: -0.08, 0.07), where the confidence intervals for the correlation coefficients where estimated using the Fisher $z$-transformation [25]. This finding on the correlation of *C. difficile* incidence across hospitals that are neighbors in the transfer network supports the use of the transfer network as a substrate for the spread of nosocomial infections.

### 2.3 Monitoring the system for hypothetical outbreaks

We used three different strategies for selecting the sensor nodes based on their properties in the static network, choosing them based on their in-degree rank, out-degree rank, or choosing them at random. Nodes with a high in-degree are expected to be efficient at funneling in pathogens from their network environment, whereas nodes with a high out-degree are expected to rapidly funnel out their pathogens.

Except for very low activation times of the order of a few days, the efficacy and the fraction of detected cases are almost unaffected by this parameter (Fig.2 left). The optimal sensor set of a strategy stabilizes after $T = 5$ days (Fig.2 right). These results corroborate that choosing sensors based on in-degree is the best overall strategy, followed by out-degree, and then the random strategy. All of the strategies result in similar sizes for the most efficient sensor sets as in the static case. In terms of the fraction of detected cases, all three strategies perform similarly, each covering about 80% of the cases. We find that the average time a sensor spends in the active state increases as a function of the activation time $T$. Therefore, an optimal approach is to choose the smallest activation time $T$ that does not deteriorate performance of the sensor system in terms of the fraction of detected cases. For an activation time $T = 5$, the average fraction of time sensors spend in the active state is 0.51 for in-degree based selection, 0.47 for out-degree based selection, and 0.46 for the random strategy.

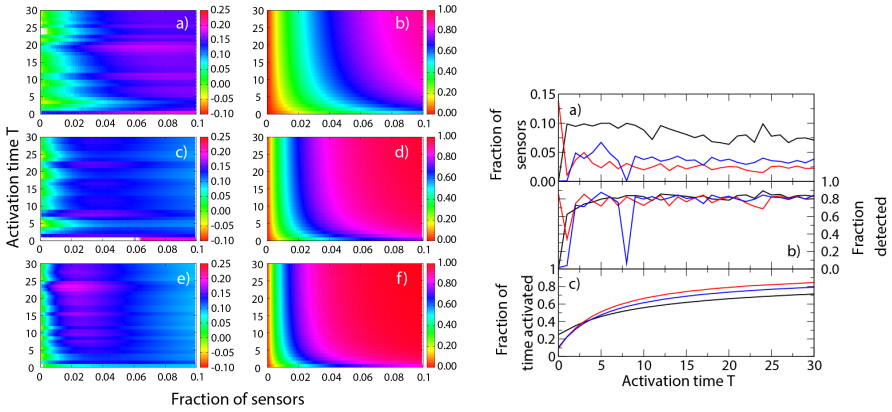


**Fig. 2. Left: The optimal sensor set.** Heatmaps showing the efficacy (left column) and fraction of detected cases (right column) on the temporal transfer network, as a function of the fraction of hospitals acting as sensors (horizontal axes) and the activity time that they implement (vertical axes). The rows of panels correspond to choosing the sensors randomly (top row), proportional to out-degree (middle row) and proportional to in-degree (bottom row). **Right: Efficacy of temporal sensor sets. a)** Fraction of sensors for the most efficient sensor set from the temporal network for sensors chosen at random (black), proportional to in-degree (red), and proportional to out-degree (blue). We have smoothened the efficacy curves by averaging the results using a window of 5 sensors. **b)** Fraction of detected cases for the most efficient sensor set. **c)** Average fraction of time that a sensor stays in the active state (same color code as on the left)

## 3 CONCLUSIONS

We studied a network defined by the transfer of 12.5M Medicare patients across 5,667 US hospitals over a 2-year period. The network is strongly geographically embedded, with 90% of all transfers spanning a distance less than 200km. The transfer network could plausibly be used as a substrate for the spread of pathogens: we observed a positive correlation for *C. difficile* incidence between hospitals and their network neighbors, identifying two qualitatively distinct regimes corresponding to low and high *C. difficile* incidence. Finally, selecting hospitals as sensors based on their in-degree in the static network was able to detect a large fraction of infections. Furthermore, an activation time of just 5 to 7 days using the dynamic sensor implementation is sufficient to achieve this surveillance with just 2% of the hospitals acting as sensors. These results support our conceptual model that the structure of the nationwide hospital patient transfer network is important for the spread of health-care associated infections, likely well beyond the illustrative case of *C. difficile* considered here. In particular, our work highlights the need to monitor the network of transfers not just individual hospitals in order to track infectious outbreaks.

Other pathogens might need a different number of sensor hospitals, a different set of sensor hospitals, or different surveillance windows. Nevertheless, the health of the entire hospital system, from the perspective of nosocomial infections or other outbreaks, could be monitored by leveraging the network structure of patient transfers.

Our study has several limitations. First, the data we used to map the hospital networks are from 2006 and 2007. However, given that hospital transfer patterns are strongly embedded in the geography of the country, as we also demonstrated here, we do not expect the age of the data to affect our results substantially. Second, we cannot assess the extent to which unobserved policies or commercial constraints might have affected the flow of patients from one hospital to another; however, these policies merely affected patient transfers, which are, in any case, observable in the current and similar future data. Third, our analyses and models assume that patient transfers are the only mechanism responsible for the spread of infections. There are, of course, other vectors or means that might result in hospitals being infected, such as the movement of physicians, nurses, and other health care staff between hospitals. Finally, in this analysis, we did not make use of the fine-scale temporal information available in transfer data; future work could evaluate how bursts of infected patients, perhaps on particular days of the week, might contribute to an epidemic.

Understanding the structure and dynamics of the hospital transfer network for the spread of real infections has a number of important implications. Empirical data could be used, either periodically or perhaps even in real time to map networks of patient movement in the US health care system, and this network could then be used monitor the spread of nosocomial and other infections in the network. In our estimation, such a system could detect 80% of *C. difficile* cases using just 2% of hospitals as network sensors. Our methods suggest practicable strategies for identifying which hospitals should serve a surveillance function for

the whole system and, in the dynamic implementation, how long the sensors should retain a higher level of alertness after each index case. These tools would be useful not only for public health interventions in the case of natural epidemics, but also in the case of deliberate ones, such as those due to a possible bioterror attack. In conclusion, the actual structure and flow pattern of patients across US hospitals confers certain specific vulnerabilities and defenses, regardless of the biology of the pathogen per se, placing theoretical bounds on any effective containment strategy directed at a contagious pathogen.

## 4 Acknowledgments

We thank Laurie Meneades for the expert assistance required to build the dataset. JFG and JPO are joint first authors of this article.

## References

1. Zimlichman E, Henderson D, Tamir O, et al. (2013) Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system. *JAMA Intern. Med.* 173(22):2039-46.
2. Threat Report 2013 — Antimicrobial Resistance — CDC. Available at: `http://www.cdc.gov/drugresistance/threat-report-2013/`
3. Klevens RM, Edwards JR, Richards CL, et al. (2007) Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep.* 122(2):160-6.
4. Infectious Diseases Society of America (IDSA) (2011) Combating Antimicrobial Resistance: Policy Recommendations to Save Lives. *Clinical Infectious Diseases* 52(S5)
5. Roberts RR, Hota B, Ahmad I, et al. (2009) Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin. Infect. Dis.* 49(8):1175-84
6. Mauldin PD, Salgado CD, Hansen IS, Durup DT, Bosso JA. (2010) Attributable hospital cost and length of stay associated with health care-associated infections caused by antibiotic-resistant gram-negative bacteria. *Antimicrob. Agents Chemother.* 54(1):109-15
7. Filice GA, Nyman JA, Lexau C, et al. (2010) Excess costs and utilization associated with methicillin resistance for patients with Staphylococcus aureus infection. *Infect. Control Hosp. Epidemiol.* 31(4):365-73.
8. Karkada UH, Adamic L a, Kahn JM, Iwashyna TJ. (2011) Limiting the spread of highly resistant hospital-acquired microorganisms via critical care transfers: a simulation study. *Intensive Care Med.* 37(10):1633-40
9. Iwashyna TJ, Christie JD, Kahn JM, Asch DA. (2009) Uncharted paths: hospital networks in critical care. *Chest*135(3):827-33
10. Iwashyna TJ, Christie JD, Moody J, Kahn JM, Asch DA. (2009) The structure of critical care transfer networks. *Med. Care* 47(7):787-93
11. Unnikrishnan KP, Patnaik D, Iwashyna TJ. (2011) Spatio-temporal Structure of US Critical Care Transfer Network. *AMIA Summits Transl. Sci. Proc.* 2011:74-78.

12. Lee BY, McGlone SM, Song Y, et al. (2011) Social network analysis of patient sharing among hospitals in Orange County, California. *Am. J. Public Health* 101(4):707-13
13. Lee BY, McGlone SM, Wong KF, et al. (2011) Modeling the spread of methicillin-resistant Staphylococcus aureus (MRSA) outbreaks throughout the hospitals in Orange County, California. *Infect. Control Hosp. Epidemiol.* 32(6):562-72
14. Huang SS, Avery TR, Song Y, et al. (2010) Quantifying interhospital patient sharing as a mechanism for infectious disease spread. *Infect. Control Hosp. Epidemiol.* 31(11):1160-9
15. Obadia T, Silhol R, Opatowski L, Temime L, Legrand J, et al. (2015) Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS Comput Biol* 11(3): e1004170.
16. Lorenzo Isella, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Rav, Caterina Rizzo, and Alberto Eugenio Tozzi. (2011) Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors *PLOS ONE* 6(2): e17144.
17. Medicare beneficiaries as a percent of total population. Available at: http://kff.org/medicare/state-indicator/medicare-beneficiaries-as-of-total-pop/
18. Overview of hospital stays in the United States, 2010. Available at: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb144.jsp
19. American Hospital Association. Available at: http://www.aha.org/
20. Gerding DN, Johnson S. Harrisons Principles of Internal Medicine. In: Fauci AS, Braunwald E, Kasper DL, et al., eds. 17th Editi. New York: McGraw-Hill; 2008.
21. Bajardi P, Barrat A, Savini L, Colizza V. (2012) Optimizing surveillance for livestock disease spreading through animal movements. *J. R. Soc. Interface* 9(76):2814-25
22. Schmiedeskamp M, Harpe S, Polk R, Oinonen M, Pakyz A. (2009) Use of International Classification of Diseases, Ninth Revision, Clinical Modification codes and medication use data to identify nosocomial Clostridium difficile infection. *Infect. Control Hosp. Epidemiol.* 30(11):1070-6
23. Scheurer DB, Hicks LS, Cook EF, Schnipper JL. (2007) Accuracy of ICD-9 coding for Clostridium difficile infections: a retrospective cohort. *Epidemiol. Infect.* 135(6):1010-3.
24. Dubberke ER, Butler AM, Yokoe DS, et al. (2010) Multicenter study of surveillance for hospital-onset Clostridium difficile infection by the use of ICD-9-CM diagnosis codes. *Infect. Control Hosp. Epidemiol.* 31(3):262-8.
25. Fisher, Ronald A (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10(4)
26. Newman MEJ, Strogatz SH, Watts DJ. (2001) Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64(2):026118
27. Newman M. (2002) Assortative Mixing in Networks. *Phys. Rev. Lett.* 89(20):208701
28. Newman M. (2003) Mixing patterns in networks. *Phys. Rev. E* 67(2):026126
29. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. (2001) Introduction To Algorithms, Chapter 16, MIT Press
30. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983) *Science* 220, pp. 671–680