

The network positions of methicillin resistant *Staphylococcus aureus* affected units in a regional healthcare system

Jan Ohst¹, Fredrik Liljeros^{2,3}, Mikael Stenhem^{4,5} and Petter Holme^{3,6,7*}

*Correspondence:

petter.holme@physics.umu.se

³Institute for Future Studies,
Stockholm, 10131, Sweden

⁶Department of Energy Science,
Sungkyunkwan University, Suwon,
440-746, Korea

⁷IceLab, Department of Physics,
Umeå University, Umeå, 90187,
Sweden

Full list of author information is
available at the end of the article

Abstract

We studied a dataset of care episodes in a regional Swedish hospital system. We followed how 2,314,477 patients moved between 8,507 units (hospital wards and outpatient clinics) over seven years. The data also included information on the date when patients tested positive with methicillin resistant *Staphylococcus aureus*. To simplify the complex flow of patients, we represented it as a network of units, where two units were connected if a patient moved from one unit to another, without visiting a third unit in between. From this network, we characterized the typical network position of units with a high prevalence of methicillin resistant *Staphylococcus aureus*, and how the patient's location in the network changed upon testing positive. On average, units with medium values of the analyzed centrality measures had the highest average prevalence. We saw a weak effect of the hospital system's response to the patient testing positive - after a positive test, the patient moved to units with a lower centrality measured as degree (i.e. number of links to other units) and in addition, the average duration of the care episodes became longer. The network of units was too random to be a strong predictor of the presence of methicillin resistant *Staphylococcus aureus* - would it be more regular, one could probably both identify and control outbreaks better. The migration of the positive patients within the healthcare system, however, helps decreasing the outbreak sizes.

Keywords: network epidemiology; methicillin resistant *Staphylococcus aureus*; hospital system; healthcare associated infections

1 Background

It has long been observed that interpersonal contacts with an ability to transmit a disease are not random as assumed by simple models of infectious disease spreading. If it is possible to estimate the characteristics of such a non-random network of contacts between individuals, we could improve the predictive and explanatory power of epidemic models. There are not so many pathogens, however, that spread over pathways where the network structure can be estimated. For this to be possible, contacts with the capacity to transmit the disease need to be discernable among all different types of inter-individual contacts, so that a network of effective contacts can be faithfully constructed. This is the case for e.g. sexually transmitted infections [1] and - the topic of this paper - healthcare associated infections (HAI) [2].

The first network-epidemiological study of the spread of disease in healthcare systems is, to our knowledge, Meyers *et al.* [3]. In this work, the authors model contagion between units populated by immobile patients. The model assumes the disease to spread between units by medical staff acting as vectors [4, 5] and is used to argue for the key-role of the staff in the spreading dynamics. Karkada *et al.* [6] and Lee *et al.* [7] make similar simulation-based studies concluding that patient transfer in critical care and nursing homes, respectively, are important factors in the dynamics of HAIs. Liljeros *et al.* [8] investigated a subset of the dataset we use in this paper. This smaller dataset recorded 295,108 inpatients from the Stockholm area of Sweden over two years. Liljeros *et al.* focused mostly on methodological questions, such as how to represent this dataset as a network of patients that is as relevant for investigating disease spreading as possible. The authors argue that different diseases need different network representations depending on their route of transmission. Ueno and Masuda [9] investigate a dataset from Tokyo community hospital sampling 388 patients and 217 doctors and nurses. They simulate disease transmission in this data and evaluate different strategies for controlling epidemics. Vanhems *et al.* [10] use a data set of similar size acquired from wearable sensors (detecting when patients or health-care workers are within a range of 1–1.5 m). They find a very heterogeneous contact structure where some health-care workers are much more central in the contact network than others. Hornbeck *et al.* [11] use a very similar data set to reach very similar conclusions. Donker *et al.* [12, 13] study a large dataset of patient flow between hospitals within the Netherlands. Their data is aggregated on a coarser level than ours - a node in the network is a hospital - but it does cover an entire nation. Donker *et al.* find a directionality of the flow towards larger, academic hospitals. This could, they argue, be exploited to control the transmission of healthcare associated pathogens (in Ref. [14] they make this point stronger by simulations and argue that just reversing the patient flow would reduce the HAI prevalence dramatically). The final network-epidemiological study of HAI we are aware of is Walker *et al.*'s study of *Clostridium difficile* in inpatients of the Oxfordshire region of the United Kingdom [15]. In this paper, the authors retrace possible transmission trees among 1,282 positive cases. They find that about 25% of the cases can be explained by an infection within the hospital system.

Currently researchers have, as seen above, either studied smaller, high precision data recorded by electronic sensors or large-scale patient referral data. These two types of data have their pros and cons - with high precision data could perhaps identify singular infection events, on the other hand, an epidemic outbreak is a large-scale phenomenon that is affected by the large-scale contact structure that at present can only be studied by patient referral data. The present paper investigated a dataset of the large-scale category.^a We use a record of all care episodes in the Stockholm region, making it possible to map the patient flow between units (that could be either a hospital ward or an outpatient clinic), we also knew who tested positive with methicillin-resistant *Staphylococcus aureus* (MRSA) - an important nosocomial pathogen - and when they tested positive. However, we did not (like Ueno and Masuda [9]) have records of the movement of the medical staff. We had to assume that the transmission of MRSA could take place outside the dataset (i.e. a patient could be infected in the community outside the healthcare system). One interesting question is how to infer these missing chains, which implicitly would mean how one can predict the false negative patients within the records of the regional healthcare system. For our data, and the methods we can envision, this would give too uncertain results at an

individual level. We would have to aggregate the results to make meaningful observations. In this work, we do not take such an individual-level approach and integrate the results. Rather, we study the system at an intermediate level - the level of health-care units. We represented the hospital system as network of units. Briefly stated, we linked two units A and B if a patient had care episodes in both units without having been admitted in any other unit in between. The links between units thus capture the possibility of infection spreading from one unit to another (or in terms of newly infected patients the link, or course, represents certainty). Just like the topology of the contact network can help us to better understand how the contact patterns between individuals affect disease transmission (which individuals that are most influential, how influential they are relative to the average, how a disease can most efficiently be mitigated, etc.) [1, 16, 17], a network of units can teach us about how the organization of the hospital system affects disease spreading. There has recently been a debate in the literature of the of the benefits of screening patients for MRSA (see Refs. [18, 19] and further references therein). A more cost effective alternative to screening all patients would be to, guided by analyses like the ones in this paper, focus on high-risk units.

2 Methods

2.1 The network of units

As in many countries, the Swedish public healthcare system is organized hierarchically into hospitals that are divided into departments that are divided into wards. In this work, we focus on at the lowest level - hospital wards and outpatient clinics - and consider the network of such units connected if at least one patient has been transferred from one unit to another. In total we study 8,507 units and 66,527,638 care episodes involving 2,314,477 individuals observed for 3,059 days. We represent this system as a network by considering a unit as a node and connecting two nodes if they, at some point in the data, had a patient transferred between each other. The links can be weighted by the number of patients transferred along it, or directed, indicating the net flow - unless otherwise stated we use the simplest representation where a link indicates the presence or absence of any patient transfer.

It is not completely trivial how to define such a transfer, especially for patients that go out of the healthcare system and then come back. The simplest solutions to this problem are either to omit the stay outside the healthcare system (and put a link between the unit that the patient is discharged from, to the first unit where the patient reenters the healthcare system), or to not add such a link at all. Since MRSA colonization can have happened before the testing, we use the first approach. The drawback is that the links no longer represent a direct referral between two units, and thus is more indirectly related to the patient flow. Another alternative approach would be to add the outside as a node, but then that node would not be easily comparable with the other nodes. For example, the real risk of transmission between individuals in that outside node would be much lower, since the probability that two persons might meet on a given day outside the healthcare system is very low.

A slight complication in our data set is that patients can be registered at different units simultaneously. It is a rather rare event (happening for about 2.7% of the patients). We represent the event that one patient is at two units a certain day by adding one unit in both directions to the weight between these units. Another feature that could affect some

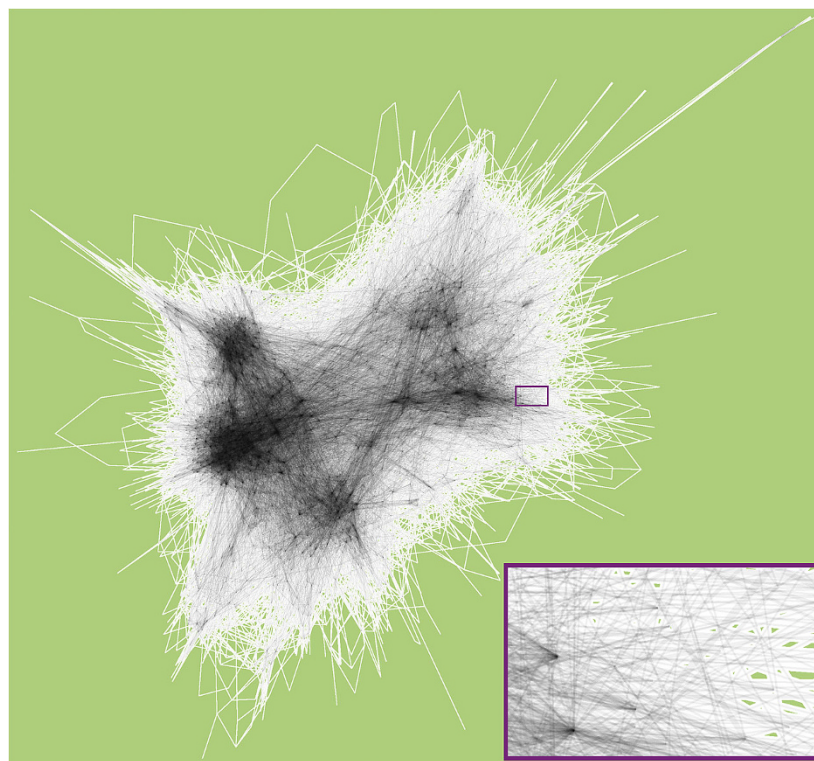


Figure 1 Ridiculogram of the patient flow. This is a visualization of the entire patient flow in the hospital system of the 8,507 units that at some point has an MRSA infected patient. It is created by a spring-embedding algorithm that forces nodes (units) exchanging many patients to be close. In this visualization, the nodes are not directly visible, only the links. The darker colors represent stronger overlap. We can see that the system is strongly connected and the hierarchical administrative organizations are not very clearly reflected in the patient flow. At the same time, there is more structure than in a purely random network which would look uniform in this type of plot - the dark blobs corresponds more or less to the major hospitals (we do not have information about what larger unit that a unit belongs to). This type of visualization should not be overinterpreted (which is often the case, hence the sobriquet “ridiculogram”). To understand the structure of the unit network, we need network metrics, which is the topic of the paper.

results is that the id number of some units (the outpatient clinics) change without the system being reorganized *per se*. Rather than cleaning the data from such short-lived nodes, we keep their presence in mind when discussing the results.

In Figure 1, we show a plot of the unit network as displayed by a spring-embedding algorithm. In this case, and unless otherwise stated, the network is aggregated over the entire sampling time. The picture that the unit network is more randomly organized than the hierarchical organization of hospital systems still holds if one plots the graph in other ways with other layout programs. Still, there are visible clusters, presumably corresponding to hospitals. In other words, the unit network has some structure that can affect MRSA transmission.

2.2 Network structural measures

We related the average prevalence of MRSA, over the study period, at unit level to different measures of network centrality. In network analysis centrality is an umbrella for a number of measures quantifying different aspects of how central a node or link is in the network [16, 17]. The following centrality measures were used:

2.2.1 Degree centralities

The simplest measures of centrality are the *in-* and *out-degree*, the number of other units that the focal unit receives patients from and transfers patients to, respectively. These measures are both local, in the sense that the centrality of a node is only affected by its neighborhood (the nodes to which it has a link). Potentially, many-step processes, where a patient is transferred through a chain of units, could be important. However, such events do not contribute to the degree centralities, which motivate the use of more elaborate metrics. If degree centrality has a capacity to explain MRSA prevalence comparable to other centrality measures, then it may even be the preferable centrality measure, since it is simple both conceptually and computationally.

2.2.2 Weighted and unweighted betweenness centrality

Another centrality concept comes from thinking about the traffic through a node. If we assume traffic originate between pairs of nodes with equal rate, and travel through the network along shortest paths, the amount of traffic through a node would then be proportional to its betweenness centrality. More technically, let $\sigma(i, j)$ be the number of shortest paths between i and j (a shortest path does not have to be unique) and let $\sigma_l(i, j)$ be the number of shortest paths between i and j that pass l , then l 's betweenness centrality is

$$C_B(l) = \sum_j \sum_{i \neq j} \frac{\sigma_l(i, j)}{\sigma(i, j)}. \quad (1)$$

This definition holds for both weighted and unweighted networks (although for weighted networks the shortest path is often unique and so the denominator is strictly one).

2.2.3 PageRank

Proposed as a measure to rank web pages, PageRank takes inspiration from a Web surfer's behavior. If people would follow hyperlinks randomly except occasionally when they go to a random page, then the popularity of a page would be proportional to its PageRank. Algorithmically, PageRank is easiest to describe as an iterative process. Let $C_P^t(i)$ be the PageRank centrality of node i at iteration t , then

$$C_P^{t+1}(i) = \frac{1-d}{N} + d \sum_{j \in \Gamma_i^{\text{in}}} \frac{C_P^t(j)}{k_j^{\text{out}}}, \quad (2)$$

where Γ_i^{in} is the set of nodes with a link pointing to i and k_j^{out} is the out-degree of j . d is a parameter that sets the balance between when the surfer follows a link and move to a random node. In this paper, we use the standard value $d = 0.85$. PageRank belongs to a class of centrality measures (including eigenvector centrality and Katz' centrality) that imagine a flow of centrality along the edges, and the actual centrality values as the steady state distribution of this flow [10].

2.2.4 Overturn

In addition to the static network measures, we also measured the overturn of patients of a unit, defined as the average number of patients entering the unit per day.

2.3 Constructing the control set

In some of our analyses, we need to compare our statistics for the MRSA-positive patients with the results for a random control set of patients that are not tested for MRSA. We generate the control set by, for every MRSA-case, finding one person from the set of non-tested patients that stays about as long time in the health care system (within 904 days) as the MRSA case. Furthermore, to get patients with similar clinical conditions, we restricted the control cases to those entering the healthcare system at the same unit as the specific patient. To make the dataset complete, we also needed to assign a test date. We choose this as the test date of the original infected person.

2.4 Prevalence

As a measure of the (relative) prevalence of MRSA at a unit, we calculate

$$P(i) = \frac{D_I(i)}{D(i)}, \quad (3)$$

where $D_I(i)$ is the total number of days a patient that has tested positive spends at unit i . $D(i)$ is the accumulated patient-days of i . A case of MRSA can be prevalent in a unit for one of two reasons. Either the case through transmission became a new case of MRSA while staying in that unit, or the case was admitted to the unit with an earlier diagnose of MRSA.

2.5 Response to infected patients

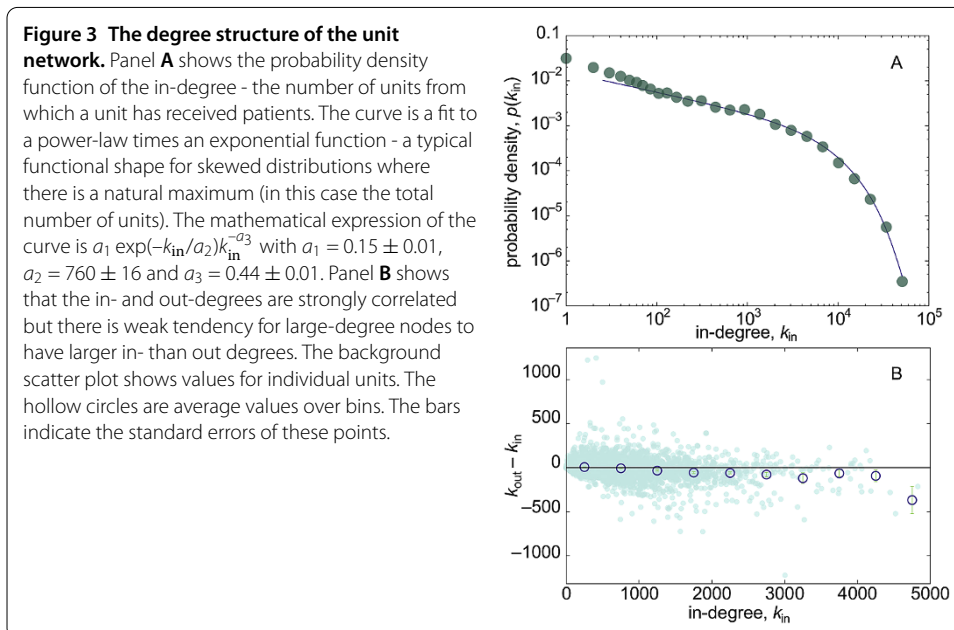
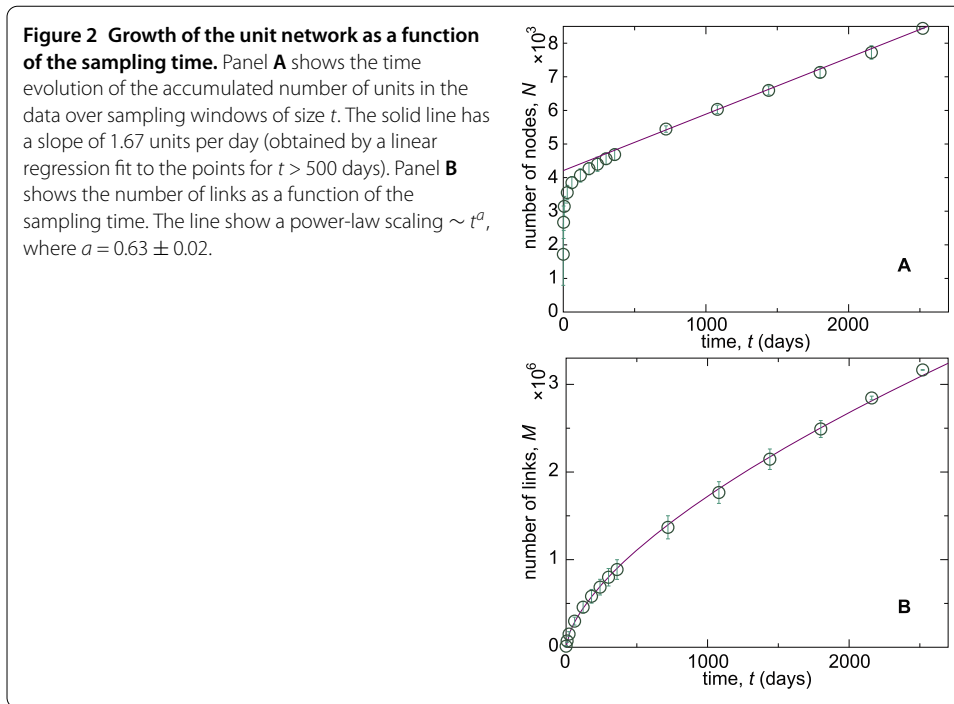
When a patient tests positive with MRSA, the healthcare system might move the patient to particular units as a precautionary measure after the diagnosis. Such units could have differing network characteristics (such as being smaller and less central). We addressed this issue by measuring network structure of the units as a function of the time when the patient was there, relative to the date of the positive test.

3 Results

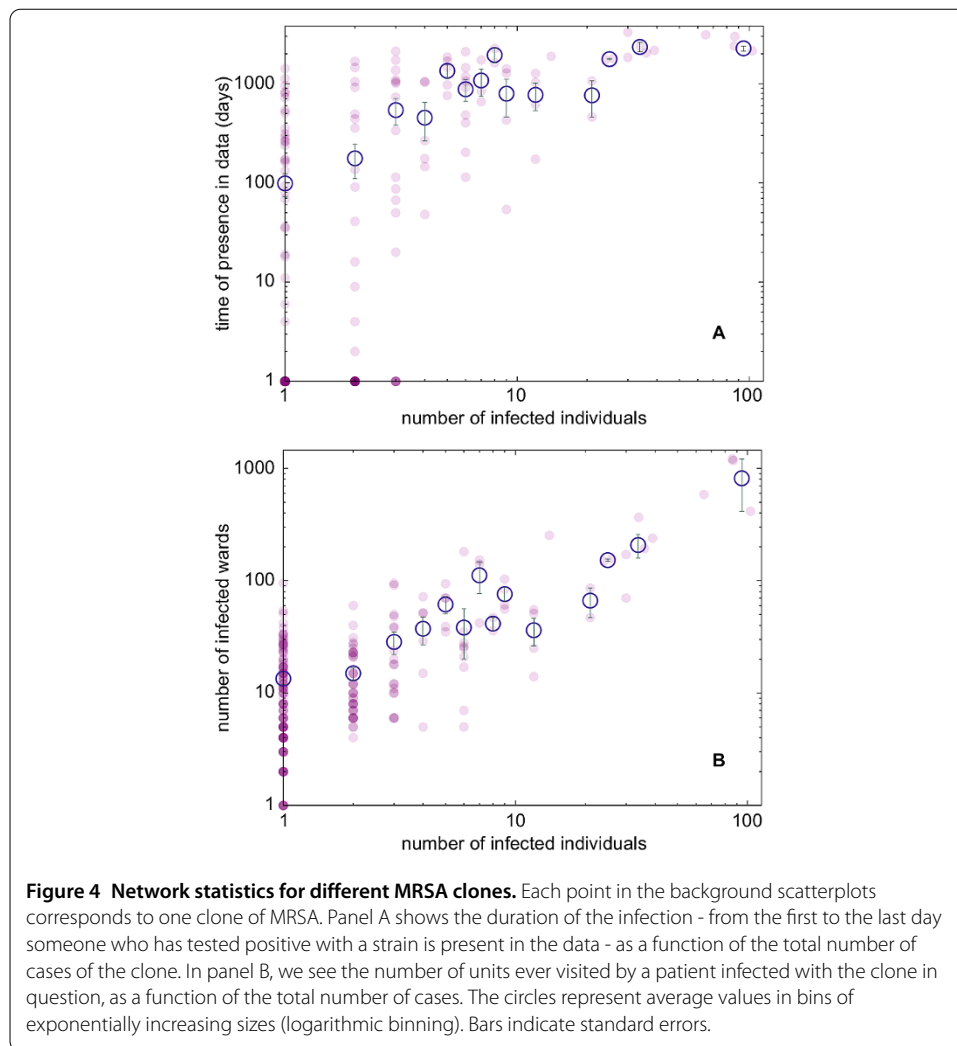
3.1 Basic structure of the network of units

The unit network had 8,507 units and 3,185,710 links giving a mean number of neighbors of 749. In Figure 2A, we study the growth of the number of units mentioned (in the context of a patient being admitted to or discharged from a unit) over windows of constant size, from random starting points. This number grew first rapidly, later following a linear increase (see Figure 2A). The rapid initial growth comes from the units present in the beginning of the dataset being mentioned (through a patient being admitted or recharged at the unit) for the first time. The later linear increase comes from reorganization and re-registration of primarily private units. The number of links between units showed a sublinear growth (Figure 2B), reflecting that the distribution of the frequency of patient transfer was broadly distributed (not shown).

Next, we tried to get a more detailed view of the network structure of the aggregated unit network. In Figure 3A, we plot its in-degree distribution - the probability mass function of the number of units that ever sent a patient to a particular unit. This distribution roughly follows power-law with an exponential cutoff. This is interesting since even broader degree distributions, like power-law distributions, make the spreading faster and epidemic thresholds lower [20]. The in- and out-degrees of units are very similar, especially in the



sense that none of the nodes with very high in-degree have a low out-degree, and vice versa (Figure 3B). This figure shows that there is a small tendency that the difference between in- and out-degree decrease with the in-degree, but the main result is that this correlation is weak. There are mechanisms that can explain both the decreasing tendency of $k_{out} - k_{in}$ and the fact that it is rather small. In the case that a patient is referred to another unit, and then returns shortly afterwards to the original unit - a common series of events - there will be a flow in both directions, the link will be bidirectional and thus contribute equally to k_{in} and k_{out} . On the other hand, if we assume some units are more in demand than oth-



ers, while at the same time not requiring more service than usual, then that would give a decreasing trend in a $k_{\text{out}} - k_{\text{in}}$ vs k_{out} plot. If there is such a celebrity effect in this data, however, we deem it too small to mention.

In summary, the unit network had a skewed degree distribution - which in principle would speed up disease spreading - but not as skewed as scale-free networks [16, 17], that has been argued to model many types of contact patterns. The network was also symmetric in the sense that the in- and out-degrees were similar between units.

3.2 Statistics for different genotypes

As mentioned, the MRSA isolates were genotyped (multi-locus sequence typing by pulsed-field gel electrophoresis [21]), which offers the possibility to tell whether MRSA transmission could have occurred between two MRSA cases that had been admitted to the same unit simultaneously when one of the cases was already diagnosed with the disease and the other case still MRSA negative. In Figure 4, we show some quantities describing the outbreak statistics of different epidemiologic types. We note from panel A that the size of the outbreaks for the different types are broadly distributed with a few types having infected around a hundred patients while the majority of clones only infected a few.

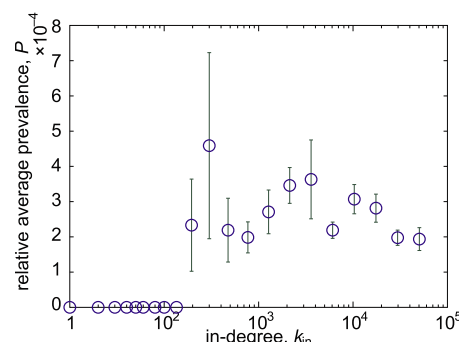
Walker *et al.* made a similar observation studying data from the Oxford region of United Kingdom [15]. Probably the more common clones are associated with infections within the health care, while the more rare ones are primarily community acquired. The duration of the presence of a clone in the data - the time between the test-dates of the first and last infected patient - varies much. The most long-lived clones are present in the data throughout the sampling period. Figure 4A shows the time of presence as a function of the total number of infected individuals of that particular clone. It is an increasing function for smaller outbreaks and seems to stabilize for longer outbreak sizes. That the points seem to converge is most likely a cut-off effect from the limited sampling time. In Figure 4B, we show the number of infected units as a function of the total size of the outbreak for all the clones. The average number of infected units scale linearly with the number of infected patients. This tells us what one would assume from the beginning - there is no difference between common and rare types in their distribution in the unit network with respect to the network position.

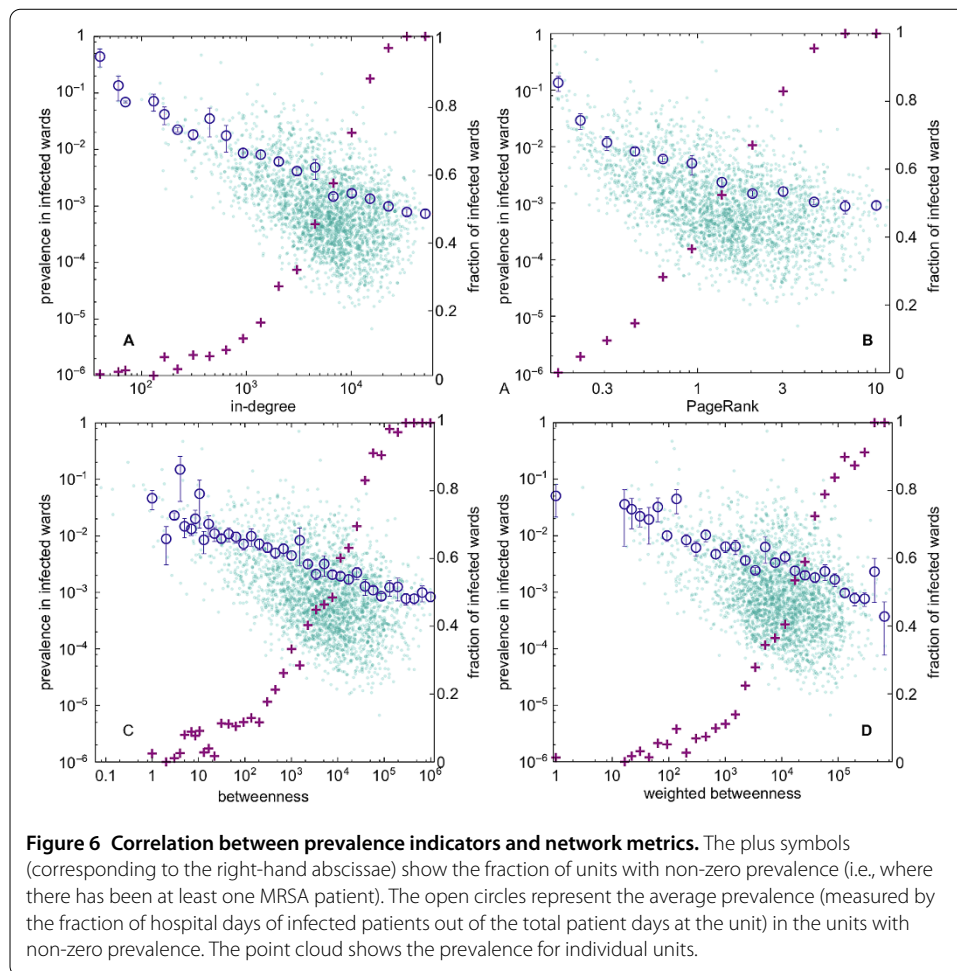
3.3 Network-structural determinants of MRSA prevalence

The most natural candidates for units that play an important role in the spreading of HAIs are the ones at the center of the network. As mentioned, there are different ways of measuring centrality, all capturing different aspects of the concept. Instead of reasoning about which one that is most appropriate in our case, we tested several of the most common centrality metrics. In Figure 5, we show the dependence of the MRSA prevalence as a function of the total number of patient days at the unit. There was a weak tendency for units of intermediate in-degree to have higher MRSA prevalence, but the chance to find a MRSA positive patient in a high-degree unit was almost as high. The low-degree units with zero prevalence are so few that, by stochastic fluctuations, they just happen to be zero. This is not remarkable - these units not only have low-degree, they had few patients too, and in the entire dataset there are only about 0.4% positive cases.

In Figure 6, we separated units with zero and non-zero prevalence and plot prevalence in the non-zero units as a function of the four centrality measures - in-degree (which according to the results shown in Figure 2 amounts to the same figure as out-degree), PageRank, betweenness and weighted betweenness. The relationship between MRSA-prevalence and the four measures of centrality was similar, as is evident from Figure 6. First, the fraction of units with at least one MRSA cases increased with a sigmoidal dependence of the logarithm of all the centrality measures. Second, of the units which ever cared for MRSA-cases, there was a negative correlation with centrality. In other words, out of the units that

Figure 5 Average MRSA prevalence as a function of in-degree. The prevalence values are averaged for all units of a certain in-degree. Prevalence is defined as the ration of patient hours of MRSA infected patients and all patients. The bars indicate standard errors. The data is logarithmically binned.

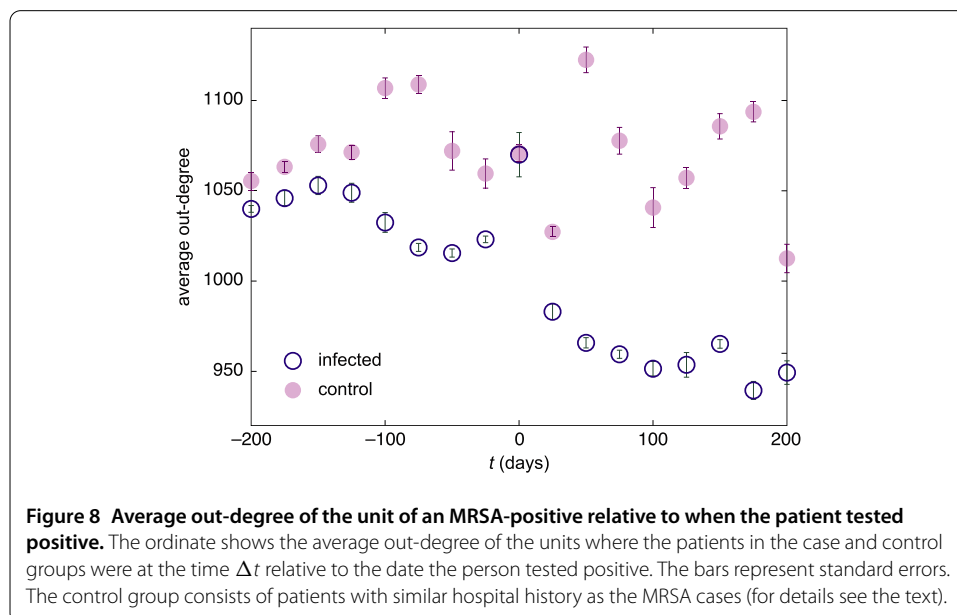
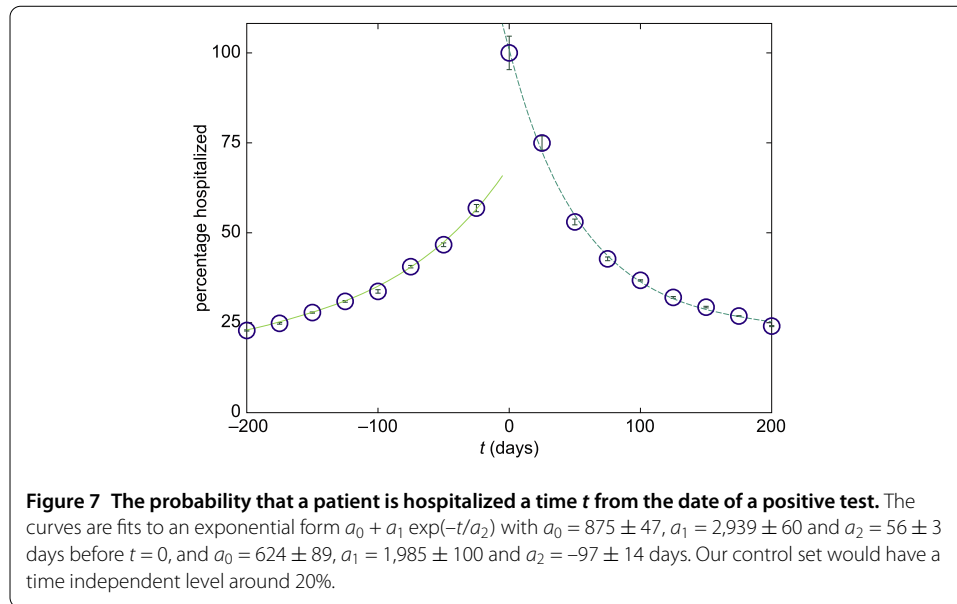




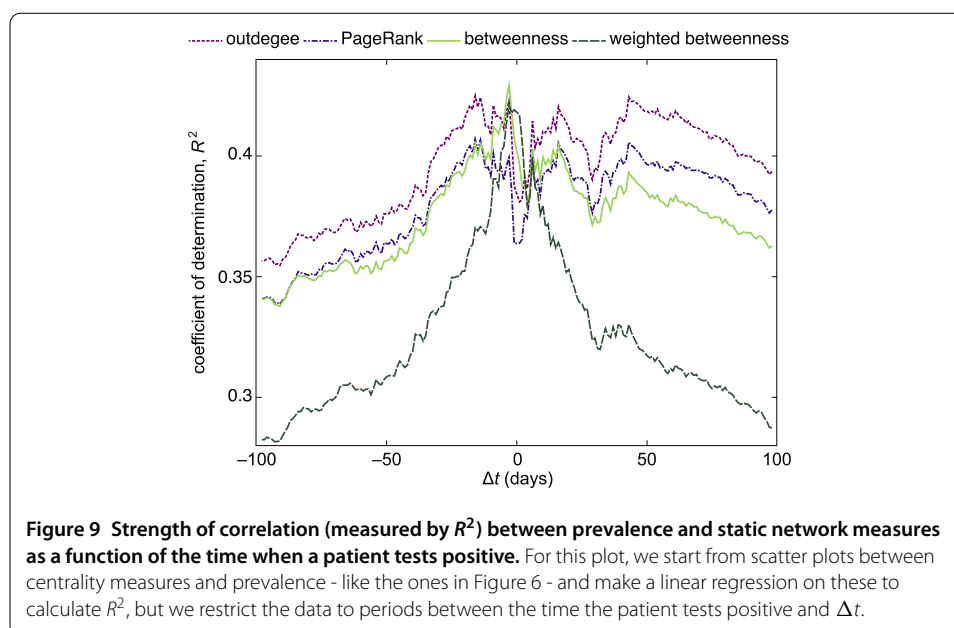
had any MRSA patient the peripheral ones (those low centrality) had a higher average load. These two factors combined did almost, but not completely, cancel out, so that the average prevalence has a weak peak for intermediate centrality values (cf. Figure 4). In summary, the MRSA prevalence has a weak dependence on centrality measures, probably too weak to be of practical use in the control of MRSA. Of course, it is often neither convenient nor feasible to impose a strict control of the patient flow, but if it followed the hierarchical, administrative organization of the healthcare systems, then the network structure would probably be more useful for controlling the disease spreading. Since that would compartmentalize the patient flow, it would also slow down the spreading - one outbreak could be effectively confined to one unit, e.g. hospital, of the healthcare system. To some extent, this is an ongoing effort within the administration of the healthcare system.

3.4 The trajectory of patients in the healthcare system

In Figure 7, we plotted the fraction of patients that were hospitalized as a function of when they tested positive t . We notice that both the increase before the test date, and decay after, is exponential. This reflects an exponential distribution of hospitalization times. It is common that patients undergo a thorough examination, including MRSA testing, on the day they become hospitalized. This explains the most asymmetric feature between the curves before and after t - the jump as t approaches zero from below.

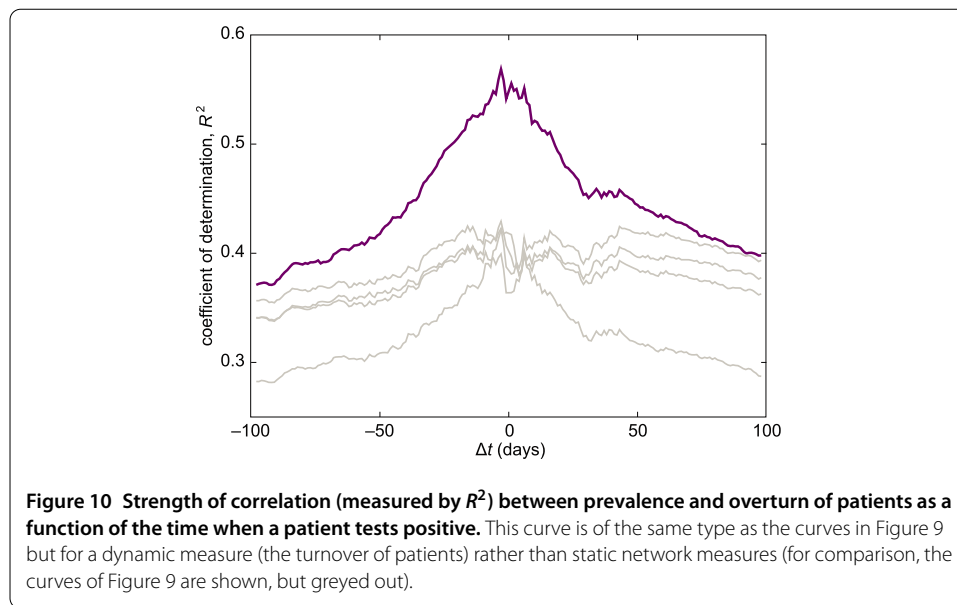


In Figure 8, we show the centrality of the unit as a function of the time relative to the date for testing positive with MRSA. The control case curve is, as predicted, quite constant. This is not true for the MRSA positives, whose units decrease in centrality with time. There can be two explanations for this phenomenon. First, this observation may reflect the response of the hospital system; i.e. to send the patients to more specialized (perhaps deliberately isolated) units. Second, it could be that the disease itself (i.e. the condition of the patient), leads the patient to units of lower centrality. From the data we have, these two scenarios are indistinguishable. This figure hints that there are more regularities to be discovered if we change from a unit to an individual perspective, something we plan for future work.



3.5 Correlations between simulated outbreak sizes and centrality

One explanation of the divergence of Figure 8 could be that the health care system was required to take precautionary measures if one treats an MRSA-positive patient. To investigate this further, we simulated disease transmission originating from one focal patient that we assumed was infective from a time $-t$ before the test date (i.e. we use the minus sign to indicate the infective period starts before the test date) to the test occasion. We used the SIR model for the disease dynamics as detailed above. In general, nodes that are more central give larger outbreak sizes. We quantify this trend by the coefficient of determination R^2 between the average outbreak size in the simulation and various measures of the centrality of the unit where the patient tested positive. R^2 can also be interpreted as a measure of the strength of the descriptive or predictive power of the centrality measure. We diagram both positive and negative t values to test the scenario that proactive measures are ineffective, or have a delayed effect. The results, plotted in Figure 9, show that the out-degree has the strongest predictive power for almost all t values (the in-degree gives a very similar curve and is thus not shown). This means that degree is the best static measure to identify risk units. The fact that degree, as a measure of centrality, discards secondary transmission events (the network two steps away from a node does not matter) suggests that only the local surroundings of the units matter in the outbreak dynamics. If one looks at the turnover of patients in the unit instead of the degree, or any other static network measure, the predictability increases much (Figure 10), which suggest that the turnover of patients is more important than the topology of the static unit network. The general, peaked shape of the R^2 vs t curves reflects that the more one includes of the patients history far from (before or after) the location at the test date, the less is the outbreak size correlated with the network structure of the unit where the patient is at the test date. This is natural, of course, but it does show that the structure of the unit network is not completely random - there is structure enough to affect the prevalence of MRSA.



4 Discussion

Since it is possible to estimate the contact structure, in terms of both time and network topology, behind the transmission of hospital-acquired disease, such diseases are well suited for studying with network theory. In this work, we analyzed a large dataset of patient flow over seven years in a healthcare system. This is such a large dataset that unless one wants to be restricted to the fastest quantities to calculate, one needs to reduce it further. One natural such reduction, the one we are investigating in this work, is to investigate the unit network (where two units are connected if a patient has transferred from one unit to another).

Just like the network of patients in close enough proximity for MRSA transmission, the unit network is not static. Indeed, private clinics can change their id numbers in the data. This phenomenon gives, effectively speaking noise to our measurements. With more consistent information about which units that split and merge, or change id number, we could model the system more accurately. Our results do still give a lower bound of the structural effects of the patient flow. Another option would be to break the network into shorter time segments during which the set of units is more stable (cf. Ref. [22]), but those segments cannot be too short - then they would not cover the infrequent links that could be very important for the size of an outbreak [23]. The network structure of the unit network is characterized by a skewed distribution of in- and out-degrees, but far from as broad distributions as power-laws (that are known to have low epidemic thresholds [16, 17]). The in- and out-degrees are strikingly symmetric, mostly because of a large fraction of reciprocal links.

Measuring the prevalence of MRSA by the ratio of patient-hours by patients that has tested positive with MRSA at the unit to the total patient hours at the unit, we conclude that there was a weak tendency for heightened prevalence for units of intermediate centrality. We also noted that the various centrality measures gave qualitatively similar results. Even though in most network models and empirical networks various centrality measures are usually positively correlated, some types of regularities can cause them to be less so, our unit network did not show any such effects. In sum, even though the healthcare system

is hierarchically organized, the patient flow in our dataset is rather random. This makes the unit network inefficient in predicting units of increased MRSA prevalence. Another reason for the weak correlations is that Sweden in general [24], and this data set in particular, has a low MRSA prevalence. This suggests that most cases are community acquired (Ref. [24] argues that most Swedish MRSA cases are infected abroad). On the other hand, around the time of the test, the MRSA carriers show a rather clear tendency to move to units that are more peripheral. Another trend we observed was that the prevalences of the different types were correlated with the centrality of the unit where the patient tested positive. This correlation was strongest when the turnover of patients was used as a (dynamic) measure of centrality. We also found that patients have an exponentially increasing probability to be present in the healthcare before the date of testing positive, and a decreasing probability afterwards. These probabilities are asymmetric in time with a larger chance of being present in the health care system after the test date. The increasing presence before the test date suggests most of the contagion has occurred within the healthcare system. The increasing presence after the test date indicates that the patients' hospitalization is related to the MRSA infection.

5 Conclusions

Although there are correlations between prevalence and centrality of units, these were too weak to be practical for identifying risk units. This could probably be changed with a more structured flow, which would also restrict the outbreak sizes (cf. Ref. [14]). The trajectory of patients shows that the disease itself and the health care's response to it makes patients move to less central units, where the expected size of outbreaks they could cause is smaller.

The fact that the more dynamic aspects of our study - both the trajectories of the MRSA-positive patients and the fact that flow is the most predictive centrality measure for the outbreak sizes - showed clearer deviations from the expected results, suggests that dynamic representations of the patient flow at a unit level could be a fruitful direction for future studies. It would also be interesting to remake the analysis with more exact data - e.g. a large-scale study of people's proximity by RFID sensors accompanied by uniform and comprehensive testing of all the patients. This would probably give clearer correlations, and also results directly derived from measurable properties of the contagion process and contact patterns.

List of abbreviations used

MRSA, methicillin resistant *Staphylococcus aureus*; HAI, healthcare associated infections.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JO, FL and PH conceived the study. JO and FL performed the analysis. JO, PH and MS wrote the manuscript.

Author details

¹Department of Mathematics, Universität Koblenz-Landau, Koblenz, 56070, Germany. ²Department of Sociology, Stockholm University, Stockholm, 10691, Sweden. ³Institute for Future Studies, Stockholm, 10131, Sweden. ⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ⁵Department of Communicable Disease Control, Stockholm County Council, Stockholm, Sweden. ⁶Department of Energy Science, Sungkyunkwan University, Suwon, 440-746, Korea. ⁷IceLab, Department of Physics, Umeå University, Umeå, 90187, Sweden.

Authors' information

This paper is a very interdisciplinary collaboration between an applied mathematician (JO), a sociologist (FL), a medical scientist (MS) and a physicist (PH). For further information, please contact the corresponding author.

Acknowledgements

The authors thank Martin Rosvall for comments. FL was supported by Riksbankens Jubileumsfond (The Bank of Sweden Tercentenary Foundation) Grant nr. P12-0705:1. PH was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2011947) and the Swedish Research Council.

Endnote

^a The data used was approved by the Regional Ethical Review Board in Stockholm (Record Number 2004/5:8).

Received: 24 June 2013 Accepted: 10 October 2014 Published online: 22 October 2014

References

1. Liljeros F, Edling CR, Amaral LAN (2003) Sexual networks: implication for the transmission of sexually transmitted infection. *Microbes Infect* 5:189-196
2. Breathnach AS (2005) Nosocomial infections. *Medicine* 33:22-26
3. Meyers LA, Newman MEJ, Martin M, Schrag S (2003) Applying network theory to epidemics: control measures for outbreaks of *Mycoplasma pneumoniae*. *Emerg Infect Dis* 9:204-210
4. Albrich WC, Halbarth S (2008) Health-care workers: source, vector, or victim of MRSA? *Lancet Infect Dis* 8:289-301
5. McBryde ES, Bradley LC, Whitby M, McElwain DLS (2004) An investigation of contact transmission of methicillin-resistant *Staphylococcus aureus*. *J Hosp Infect* 58:104-108
6. Karkada UH et al (2011) Limiting the spread of highly resistant hospital-acquired microorganisms via critical care transfers: a simulation study. *Intensive Care Med* 37:1633-1640
7. Lee BY et al (2013) The importance of nursing homes in the spread of methicillin-resistant *Staphylococcus aureus* (MRSA) among hospitals. *Med Care* 51:205-215
8. Liljeros F, Giesecke J, Holme P (2007) The contact network of inpatients in a regional health care system: a longitudinal case study. *Math Popul Stud* 14:269-284
9. Ueno T, Masuda N (2008) Controlling nosocomial infection based on structure of hospital social networks. *J Theor Biol* 254:655-666
10. Vanhems P et al (2013) Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* 8:e73970
11. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM (2012) Using sensor networks to study the effect of peripartetic healthcare workers on the spread of hospital-associated infections. *J Infect Dis* 206:1549-1557
12. Donker T, Wallinga J, Grundmann H (2010) Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLoS Comput Biol* 6:e1000715
13. Donker T, Wallinga J, Slack R, Grundmann H (2012) Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PLoS ONE* 7:e35002
14. Donker T, Wallinga J, Grundmann H (2014) Dispersal of antibiotic-resistant high-risk clones by hospital networks: changing the patient direction can make all the difference. *J Hosp Infect* 86:34-41
15. Walker AS et al (2012) Characterization of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. *PLoS Med* 9:e1001172
16. Newman MEJ (2010) *Networks: an introduction*. Oxford University Press, Oxford
17. Estrada E (2011) *The structure of complex networks*. Oxford University Press, Oxford
18. Jain R et al (2011) Veterans Affairs initiative to prevent methicillin-resistant *Staphylococcus aureus* infections. *N Engl J Med* 364:1419-1430
19. Huskins WC et al (2011) Intervention to reduce transmission of resistant bacteria in intensive care. *N Engl J Med* 364:1407-1418
20. Wang B, Cao L, Suzuki H, Aihara K (2012) Safety-information-driven human mobility patterns with metapopulation epidemic dynamics. *Sci Rep* 2:887
21. Melles DC, van Leeuwen WB, Snijders SV, Horst-Kreft D, Peeters JK, Verbrugh HA, van Belkum A (2007) Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. *J Microbiol Methods* 69:371-375
22. Holme P (2013) Epidemiologically optimal static networks from temporal network data. *PLoS Comput Biol* 9:e1003142
23. Moore C, Newman MEJ (2000) Epidemics and percolation in small-world networks. *Phys Rev E* 61:5678-5682
24. Christenson B, Arndung B, Sylvan S (2011) Methicillin-resistant *Staphylococcus aureus* infections in Uppsala County, Sweden. *Open Infect Dis J* 5:107-114

doi:10.1140/epjds/s13688-014-0029-6

Cite this article as: Ohst et al.: The network positions of methicillin resistant *Staphylococcus aureus* affected units in a regional healthcare system. *EPJ Data Science* 2014 **3**:29.