IMT

SCHOOL
FOR ADVANCED
STUDIES
LUCCA

#01
2018

ISSN 2279-6894
IMT LUCCA EIC WORKING PAPER SERIES 01
January 2018

RA **Economics and institutional change**

The Global Health Networks: A Comparative Analysis of Tuberculosis, Malaria and Pneumonia Using Social Media Data

Milena Lopreite
Michelangelo Puliga
Massimo Riccaboni

Research Area
Economics and institutional change

The Global Health Networks: A Comparative Analysis of Tuberculosis, Malaria and Pneumonia Using Social Media Data

Milena Lopreite
Scuola Superiore Sant'Anna

Michelangelo Puliga
IMT School for Advanced Studies Lucca

Massimo Riccaboni
IMT School for Advanced Studies Lucca

ISSN 2279-6894
IMT LUCCA EIC WORKING PAPER SERIES #01/2018

Piazza San Ponziano 6, 55100 Lucca

The Global Health Networks: A Comparative Analysis of Tuberculosis, Malaria and Pneumonia Using Social Media Data

Milena LOPREITE[1] Michelangelo PULIGA[2] Massimo RICCABONI[3]

[1]Institute of Economics - Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33, 56127 Pisa, Italy. Email: *milena.lopreite@santannapisa.it*

[2]IMT-Scuola IMT Alti Studi Lucca, Piazza San Francesco 19, 55100 Lucca, Italy; Linkalab, Complex Networks Computational laboratory.

[3]IMT-Scuola IMT Alti Studi Lucca, Piazza San Francesco 19, 55100 Lucca, Italy;

Abstract

Global health networks (GHNs) of organizations fighting major health threats represent a useful strategy to respond to the challenge of mobilizing and coordinating different types of health organizations across borders toward a common goal. In this paper we reconstruct the GHNs of malaria, tuberculosis and pneumonia by creating a new unique database of health organizations from the official Twitter accounts of each organization. We use a majority voter Multi Naive Bayes classifier to discover, among the Twitter users, the ones that represent organizations or groups active in each disease area. We perform a social network analysis (SNA) of the global health networks (GHNs) to evaluate the structure of the network and the role and performance of the organizations in each network. We find evidence that the GHN of malaria, TBC and pneumonia are different in terms of performance and leadership, geographical coverage as well as Twitter popularity. Our analysis validate the use of social media to analyze GHNs, their effectiveness and to mobilize the global community toward global sustainable development.

Keywords: global health network; social network analysis; machine learning classifier; tuberculosis; malaria; pneumonia; policy evaluation

JEL: i15, i18, c8

1 Introduction

Effective health plans contribute to increase development by improving quality of life and by ensuring social and territorial cohesion (Fogel, 2004). While developed countries spend a high proportion of their GDP in health care and guarantee universal health coverage, the developing countries are still lacking of domestic resources to cope with their needs. In 2015, the United Nations have identified health and well-being as one of the 17 sustainable development goals. Key priorities are related to child and maternity health, as well as prevention and treatment of the main infectious diseases, such as HIV, malaria and tuberculosis (TBC). The main international organizations such as WHO and UNICEF are to increase the amount of aid provided to less developed countries. Their aim is to help the developing countries to reduce the financing constraints by directly increasing the financial assistance in health and by improving the health workforce in the least developed countries.

Donors and financing organizations need not only to increase the aid volumes but also to provide a more effective aid (Chong, 2002; Emmerij, 2014; Kragelund, 2011; McEwan and Mawdsley, 2012; Pronk, 2003; White and L.Woestman, 1994; White and McGillivray, 1995). The aid fragmentation and lack of coverage are significant problem for the effective development of the cooperation. In the health sector, characterized by many different players, each one with different approaches and priorities, cooperation should be developed despite high transaction costs (Acharya et al., 2006; Annen and Kosempel, 2009; Gehring et al., 2017; Kimura and Sawada, 2012; Munro, 2005). For this reason the outcome is still mixed especially in the poor countries where the majority of financing for the health activities derives from foreign aid. Even though there has been some progress in the M. East, N. Africa and S. Asia, the child mortality is still too high (WHO, 2007).

It is in this context that the creation of global health networks (GHNs) - involving all the actors devoted to a specific disease or to a class of related diseases - is an instrument that can help the international community to meet health sustainable development goals by 2030. Furthermore, GHNs play a fundamental role to rapidly mobilize the international community to face urgent and unforeseen challenges such as major disasters and the outbreaks of new infectious diseases (Ferguson et al., 1995).

The networks of health organizations collaborating on an international scale are called global health networks (LaPorte, 1994; Shiffman et al., 2016). Generally speaking, the main goal of a global health network (GHN) is to improve the population health. However there are some health networks that are better than others in terms of advocating support, attracting funding, mobilizing communities and obtaining policies interventions. Adequate governance and leadership are needed to ensure that each actor is in charge of a task that contributes to the efficacy of the network. Governance, coordination and leadership are indeed among the most important issues to be solved in the organization of GHNs (Shiffman et al., 2016).

Currently, some of the main areas of intervention in global health are: HIV, reproductive health, child survival, malaria, tuberculosis (TBC) and pneumonia (Berlan, 2016; Firestone et al., 2017; Quissell and Walt, 2016; Shiffman et al., 2016). Despite the recent progress and the millions of lives saved, global actions fall far short in many respects.

In this study we focus on TBC, malaria and pneumonia. TBC is one of the major causes of death in the world as it provokes more death with respect to malaria, pneumonia and HIV (WHO, 2016). According to the Global Tuberculosis Report of World Health Organization (WHO, 2016) between year 2000 and 2015 the TBC treatment globally saved 49 millions of lives. In 2014 the spending by governments on health were behind the WHO benchmark of at least 6 per cent of GDP in 150 countries. It is estimated that in low and middle-income countries are necessary for TBC treatment almost 8.3 USD billion per year; 6.6 USD billion were available in 2016 in countries for the TBC response and they were funded respectively for the 84 per cent by the domestic sector and for the 16 per cent by the international donor financing. Also if a large amount of resources was used for health policy interventions, the Global Tuberculosis report underlines that there is a funding gap of 1.7USD billion in investments for TBC prevention, diagnosis and treatment. Moreover, at least one more USD billion per year is necessary for research and development. In the case of malaria the progress towards global targets is behind the ones for TBC. In fact, between 2010 and 2015 there was a global reduction of the malaria incidence rates of 21 per cent, in the African Region even if the governments provided only the 32 per cent of total funding to contrast the malaria burden (World Malaria Report 2016). The United States of America and the United Kingdom were the largest international contributors with 35 per cent and 16 per cent of the total funding, respectively (WHO, 2016). Finally, the pneumonia disease represents one the major cause of death for children under five years of age by killing around 920,000 children in 2015 (WHO, 2016). Although the implementation of effective interventions plans has reduced pneumonia mortality from 4 millions in 1981 to just over one million in 2013, pneumonia still accounts for nearly one-fifth of childhood deaths worldwide (WHO, 2016). It is most prevalent in south Asia and sub-saharan Africa with respectively 13 per cent and 7 per cent of death rates due to pneumonia under five years old.

Many Global Health partnerships (GHPs) such as The Global Fund for AIDS, Tuberculosis and Malaria, are working to fight against these diseases. The proliferation of GHPs has created, or exacerbated, new problems such as poor coordination and function duplication among different donors, a higher administrative burden, and misalignments of goals and actions across organizational and national borders.

With this in mind our analysis starts from the pioneering study of Shiffman et al. (2016) about GHNs from which we introduce several novelties. First, with respect to Shiffman et al. (2016) that is more interested in a qualitative approach to the GHN framework, we instead propose a quantitative study using standard network measures. In this vein, our paper it is not only a simple extension of sociological studies on GHN but it represents the first tentative to investigate the relation between social network analysis, GHN and policy interventions in terms of donors, organizations and the effective development of the cooperation in developing countries. We introduce several metrics from Social Networks Analysis that can be used to measure the structural properties of GHNs to predict their performances from the global to the local scale.

Second, to accomplish the analysis, we build a new and unique database of health organizations related to each disease exploiting all social information from the official accounts on the Twitter platform. The creation of this database uses modern machine learning classification technologies and can be replicated for other diseases as well.

We collect the actual composition, size and structure of a global health network for TBC, malaria and pneumonia. Once extracted the GHNs, we perform a social network analysis:

a) to compare each GHN using global network measures such as size, density, focus, overlap and average centrality indicators; b) to analyze the effectiveness of each organization in fighting TBC, pneumonia and malaria diseases in low and middle-income countries by using local network indicators of centrality (Jablin and Putman, 2001) like betweenness, closeness and degree; c) to study the leadership within each global network; d) to study the geographical distribution of the nodes with respect to the spread of the disease (global coverage).

Finally, one of the most interesting parts of this analysis is the creation of a set of practices, and methodologies aiming at empowering the non-governmental organizations (NGOs) and official health organizations with tools to evaluate both the quality of the intervention (in terms of media response and people's engagement) and the actual composition of those global health networks (or the equivalent aid networks).

Discovering the actors to the level of individuals proved to be a complex task, measuring response is the key to improve the engagement. Several suggestions on how to effectively use Social Networks tools for this goal were already in literature and in the NGO practices, for instance Trevor Thrall A. (2014) discuss how the NGO can influence the press and obtain success despite the challenges in communication that were not solved by the social media platforms and by the internet.

The NGO Save the Children published an handbook of network analysis (Dershem, 2011) showing an example of the usage of on-field surveys both to understand who are the partners and the local actors and how to measure the relative influence of the participants. In our study, we take this work one step further using Twitter as a main source of node discovery and performance measurement. It is the diffusion of the social network that makes the methodology a promising technique to collect and measure the effectiveness of the actions.

The remainder of this paper is organized as follows. In section 2 we present the social network analysis and we focus on the main network indicators (local and global network measures), we evaluate the effectiveness of the organizations and we compare GHNs. We also introduce the concept of global aid in health and of GHN partnerships by including the health network studies related to TBC, malaria and pneumonia. In section 3 we briefly describe the data and the methodology of the network analysis. In section 4 we discuss the results in terms of global SNA measures (size, density, focus, overlap), local SNA measures (leadership, Twitter popularity) and geographical coverage. Finally, in the last section we discuss the policy implications of the social network analysis and how this analysis could help practitioners and policy makers in reaching new users, engaging activists and eventually improving fundraising, the organization and effectiveness of GHNs.

2 Network analysis

2.1 Global health networks and development

The GHN is defined as a set of interconnected nodes represented by individuals, groups, organizations that at a country level, or internationally share information, organize campaigns and collaborate to solve a particular global health problem.

A key goal of many organizations active in the health sector is to improve the effectiveness of aid in terms of development assistance and policy harmonization. In fact health policies,health systems and the instruments that support the aid are designed to maximize the impact of better health on poverty, and to fulfill the needs of poor people.

The effectiveness of the entire global health network depends, in general, on four elements (Hafner-Burton et al., 2009; Isett et al., 2011; Kahler, 2009; Lecy et al., 2014; Sikkink, 2009): leadership, governance, composition and framing strategies (Shiffman et al., 2016). A good leadership is important to pursue a given objective. An appropriate governance facilitates collective actions. An heterogeneous network increases the cohesion of its components and it obtains better outcomes than an uniform one (Hong and Page, 2004; Page, 2007). According to the framing strategy the networks are likely to be more effective if the actors attract attention and obtain more resources with public positions or media campaigns. Finally, the performance of the networks increases with more funding thanks to unrestricted norms and to the presence of more liberal groups that facilitate their expansion.

Many studies analyze the effectiveness of the network by considering outputs, impact, policy harmonization and, policy consequences (Sabatier, 2007; Shiffman et al., 2016; Weiss, 1972; Wholey, 1983).

The policy outcomes are very often influenced by other variables that may be important in different contexts such as policy environment, other related diseases, the effort of each actor in the network, etc.

For example the emergence of HIV/AIDS led to a renewed attention of TBC in W. Europe and N. America in the 1980s. In several cases the network facilitates collective actions while in other cases the coalitions are weak or even not existing. The TBC control strategy recommended by the World Health Organization called DOTS (directly observed treatment, short-course) was highly useful for the most affected countries (WHO, 2016). The Global Fund (https://www.theglobalfund.org/) fighting malaria, HIV and TBC is well coordinated and goal-oriented. They positively influenced cross national policies and their actions resulted in an increase of health interventions. Instead, the pneumonia network between 1981-2013 contributed in the African Region to a decline in disease by only 3 per cent and this underlines a limited influence on international policies (WHO, 2016). Other differences in the effectiveness of TBC, malaria and pneumonia networks are observed in terms of framing strategies and political composition (Berlan, 2016; Quissell and Walt, 2016). Especially in the last years the strong coalition of researchers, donors and organizations in the TBC and malaria network attracted attention and increased resources by influencing health policy plans. In contrast the pneumonia network was fragmented with low coalition and decentralized communities of individuals and organizations. Its structure of connectivity was unable of efficient information sharing, ideas and resources. Its expansion and growth over time was limited with respect to the TBC and malaria network. In this case the pneumonia network represents a force of secondary importance in obtaining global agreements and leadership development intervention. The main differences between the pneumonia network and the other networks is the absence of powerful and specialized organizations such as RollBackMalaria or IAVI for TBC: all interventions have to be carried out by the major official organizations such as WHO or Global Fund.

Finally, the focus toward the harmonization is primarily on how donors can better support public policies. As in many low-income countries a large share of both health financing and provision occurs in the private sector. For this reason the political agenda has to include ways in which governments can build networks with civil society, NGO groups and private business to achieve an effective harmonization. The search for new resources, and the need to attract new donors, has led to the creation of new aid partnerships, such as the Global Fund to Fight AIDS,Tuberculosis and Malaria, the International Finance Facility and the Millennium Challenge Account. The challenge is how to increase the funding level without increasing the administrative burden.

3 Methods

3.1 Twitter and global health organizations

The Twitter micro-blogging platform is a social network with peculiar rules of friendship. Links among its users are distinguished in *follower link* (link from a node to the others) and in *friendship link* (link from the others to the node). This asymmetry plays an important role in the platform as it identifies per se the *strong ties* of the friendship and the *weak ties* of the follower relationship. Probably the most important property of Twitter for social scientists is its openness to the social graph exploration. Twitter is the only major social network that allows to freely navigate and extract the network structure attached to each user. While, for instance Facebook, Instagram and Youtube impose strict privacy rules. Furthermore Twitter has no upper limit of how many friends/followers a user can have.

Recently Twitter data and machine learning techniques have been used to nowcast the evolution of socio-economic systems. For instance, Twitter has been used to estimate the sentiment of users about stocks on the market: if a stock is heavily commented in the social media it is likely to be volatile on the market (Bollen et al., 2011). Another study shows how Twitter data can be successfully used to estimate the unemployment rate in US, correlating usage patterns and unemployment levels at a province level (Bokányi, 2017).

Social media data can also be used to reconstruct business relationships. Zhongming M. (2011) use Machine Learning techniques from online news as a mean to reconstruct the competitor and partnership relationships of a group of companies.

The openness of the platform makes Twitter a tool to discover official accounts of companies or organizations that, usually, have *media managers* to follow the account activities. Media managers are responsible to talk with the public, receiving questions and complaints. Moreover they interact with other companies forming connections that can express collaboration, business affinity or even competition.

In the case of the GHN we use the official accounts of the organization related to each disease [1]. Our basic hypothesis is that if two organizations A and B are connected by a friendship relationship in Twitter then it is likely they will collaborate or at least they endorse each other.

When screening the Twitter friends of an organization we need to recognize, among the users, the ones that are organizations or groups having an interest in the disease. To accomplish this task we need a *classifier* that can guess automatically which friend is an organization. Since the majority of the organizations in Twitter are acting to a global level (as from their description and their language) we can guess the actual composition of a GHN and use the SNA to study its features.

Using Twitter we can also study the success of the media campaigns (if any) of each organization. The way they gained users reveals the efficacy of their public initiatives and the strength of their communication.

3.2 Creating the Twitter GHN of a disease with machine learning techniques

The main hurdles in reconstructing the GHN are the global coverage of the data collection procedure and the transnational nature of the network. Social media, like Facebook and Twitter have a global reach, an unbiased coverage across the north-south divide, and are natural candidates to be considered as key sources of information about the activity of health organizations and their relationships. Social media are natural sources of information about inter-organizational ties since they provide detailed and timely information about different social ties. We choose to use Twitter data in this study due to network traversal restrictions in Facebook, i.e. the impossibility of navigating a friend list of a user. Twitter has more than 350 million users on a global scale with a better coverage of English speaking countries(Kulshrestha et al., 2012). We extend the coverage of institutions across linguistic groups collecting Twitter accounts in English, French and Spanish. Those three languages combined are either first or second language of choice in most of the countries with a few notable exceptions like the Balkan countries which, by the

[1]The organizations that belong to a GHN are research institutions, charities, official UN organizations such as the WHO and other groups. We use the name *organization* for all members of the network with no role distinction

way are marginal, in our study (Mocanu et al., 2013). Since Twitter does not provide an uniform global coverage, countries such as China and the Arabic world remain underrepresented in our analysis. However as Twitter has recently become a privileged media to promote the activities of health organizations and diffuse information to patients, donors and the general public it remains one of the most affordable and reliable tool for media campaign to attract financing and support(Jones, 2017; Swamy et al., 2017).

Our database consists of official Twitter accounts of health organization connected by friendship relationships. The creation procedure involves an initial manual input on the Twitter Search field (`https://twitter.com/search`) looking for the health organizations fighting a disease. The goal is defining an initial set of health related organizations for each disease. Seed organizations are then used to generate two levels of friends starting from each organization, i.e. we explore the layer of the friends, then the layer of friends of friends (2nd degree neighbors). Each candidate friend node is then classified as health / non health organization using a pre-trained machine learning classifier using all the information provided in its profile (location, name, url and self-description). Condition to belong to this network is that organizations speak English, French or Spanish and have a profile compatible with groups fighting the disease.

The classifier uses a *majority voter* with 100 runs of the Multi Naive Bayes classifier (Bauer and Kohavi, 1999; S. et al., 2016) with a manual trained set consisting of more than a thousand candidate organizations used as training and test set. With the help of this tool we are able to recognize the official accounts of GHN members with a precision above 90 per cent. Several manual checks are done on the resulting network to inspect the nodes.

As a final improvement we enrich the network with the location of the organizations (doing some text cleaning and name regularization on the location names) as extracted from the location field of each Twitter account. User location is analyzed to study the geographical coverage and influence area of the GHNs. Notice that studying the geography of an GHN is different of studying the geographical coverage of each organization. In this study we estimate the distribution of each GHN in the world not the local reach of individual organizations belonging to a GHN.

3.3 Social network analysis

In our work, we apply Social Network Analysis (SNA) to Twitter data. SNA is a prominent tool to analyze the structure of social interactions among individuals, groups and organizations in economics, organizational studies and the social sciences (Burt, 1992; Emirbayer and Goodwin, 1994; Granovetter, 1973; Heclo, 1978; Jablin and Putman, 2001; Parkhe et al., 2006; Powell, 1990; Podolny and Page, 1998).

Different types of organizations use social network to strengthen their relationships with stakeholders. The basic elements of the network are defined as nodes while the connections between the nodes are the links. An important type of *connectivity* that help to predict the efficiency of an organization, is the bridging among nodes that indicates better collaboration among different groups resulting in opportunities for innovation and growth. For the companies, the bridging action can often increase their performance because the bridgers access to information that is in general unknown to the others member of network (Burt, 2005). Centrality measures, such as degree, closeness and betweenness centrality, have been used as proxy of power, prestige, better access to information, and leadership (Jablin and Putman, 2001). The degree is the total number of connections from other actors to the given actor and it is a good proxy of the leadership development in an organization (Freeman, 1979). The closeness is the average path distance from an actor to all other nodes. It expresses how fast an actor can reach each other actor in the network. Finally, the betweenness measures the brokerage position of an actor with respect to two or more other actors. It indicates how much the actor is able to influence communication between other actors.

The most basic structural characteristics of a network are its size and its density. The size of a network is the total numbers of nodes (such as organizations or individuals), participating in a network. The density (connectedness) is calculated as the total number of the links in a network divided by the theoretical number of possible linkages (Hanneman and Riddle, 2005). Given a network of n nodes the total possible number of links is given by the combinatorial quantity: $n(n-1)$ or $n(n-1)/2$ if the network is undirected and the links have no orientation.

At a micro level (i.e organizations and small groups of organization) the density represents the integration between the actors and it underlines the existence of possible strong relations (Bueno et al., 2004; Paxton, 1999). The organizations with strong social ties can gain a great competitive advantage or they can increase their social capital.

4 Results

The results section is organized in three parts. Firstly, we make an analysis of each GHN using global SNA measures such as size, density, focus and overlap of the topics covered by network, and average centrality indicators. Secondly, we analyze individual nodes of the network in order to reveal the leaders, the network bridgers and the most popular nodes on Twitter. Finally, we study the geographical coverage of the each network by sub-continent and by country with comparisons with official statistics of each disease and GDP per capita.

4.1 Global network measures

The size of a network is related to the diffusion and popularity of the disease and to the relationships with other diseases. For instance the TBC network is the largest one having 1,838 nodes which are partially overlapped with the HIV/AIDS network. This network received a lot of attention in the past thirty years as TBC is the first death cause in many HIV patients in the low middle-income countries, the two GHN are clearly interconnected. The malaria network is of 196 nodes while the pneumonia one has 65 nodes (Table 1). The density of the three networks is also different and inversely proportional to the size of the network: the small pneumonia network is denser with a value of 7.3 per cent, followed by the malaria with 1.7 per cent and finally by the TBC network with 0.2 per cent (Table 1). Smaller networks, usually, tend to have high density, this helps to maintain a functioning network.

Table 1 about here

The average degree of a network reveals, along with the density, the average power of its components. In Table 2 we report these statistics.

Table 2 about here

The focus of a network can be measured in those Social Networks where a user can choose an username that indicates commitment to a cause. For instance a node such as "EndMalariaNow" is clearly pointing to the malaria disease, while a user like "WHO" despite its importance is not focused on a specific disease. Counting the number of nodes having in their username the reference to the disease is a simple measure of *focus* for each network. With this indicator we can estimate malaria as the most focused network, while pneumonia is the less focused one (Table 3). The TBC case needs to be discussed in more detail. The disease is strictly related with HIV, in poor countries the HIV organizations are then actively targeting TBC in their programs. For this reason there is a relatively large overlap of the TBC network with the HIV one. To get the focus of the TBC-GHN, we need to sum up the HIV and the TBC related nodes: the network focus is then represented in the Table 3 by the column ("HIV/AIDS + TBC") and reach 6.0 per cent. As a summary of the focus measures, the most focused network is malaria with 11.7 per cent, followed by TBC with 6 per cent, while the pneumonia network has a focus of only 3.1 per cent.

Table 3 about here

We can measure the overall popularity of each GHN on Twitter with two variables: total number of tweets, and total number of followers. The two measures as well as their averages give us an indication of the activity and of the popularity of each network.

The results in Table 4 show, as expected, the large activity of the TBC network, followed by malaria and pneumonia. However the average user activity (number of tweets per user) and average user popularity (number of followers per user) is slowly rewarding the efforts of the pneumonia organization to catch up the other networks (Table 4).

Table 4 about here

4.2 Local network measures

The analysis of local network measures identifies the most active nodes, the most popular nodes on Twitter and finally the network leaders. The first three hubs (by node degree) of each network represent the most connected nodes and they are reported in Table 5.

Table 5 about here

In the case of the more mature TBC network the hubs come at lease in part from organizations fighting the HIV disease (PEPFAR). We see here that for the smallest network pneumonia the hubs are official organizations, large non governmental organizations, or charities. Finally, for malaria nodes such as RollBackMalaria, MalariaVaccine are global leaders: they are hubs devoted to the disease.

Centrality of the nodes is measured with closeness and betweeness. The most central nodes are shown in Table 5. Here we see that WorldVision is both the best bridger and the largest hub for TBC, TBAlliance is a node close to all others. RollBackMalaria is an hub and the most central node (in terms of closeness) while is MalariaVaccine that is bridging the network. Pneumonia is dominated by the gatesfoundation (hub) and by the WHO acting as a bridger.

As a general remark if a network is mature the leaders are no more official health organizations (WHO etc.) or large charities (Gates foundations etc.), but they are nodes with a clear mission (MalariaVaccine, TBAlliance etc.). Vice versa when the network is weak the leaders are the official health organizations (WHO) or the large charities as in the case of pneumonia.

Similarly to the degree and to the centrality of each node there is a Twitter specific measure of popularity: the follower to friends ratio (Gurajala et al., 2016). Popular users on Twitter have a large ratio (i.e. many more followers than friends). In our case the most popular nodes are the organizations operating in the United States often related to the media/internet domain (Table 6). For instance the Gates foundations (Melinda and William Gates) are very popular on Twitter. Unfortunately the followers to friend ratio in the study of the GHN is not completely reliable as leadership measure as not all followers are interested in the disease, and not all friends are organizations. This ratio, however, is related to the success of media campaigns.

Table 6 about here

4.3 Geographical coverage of the GHN

To study the geographical coverage and the relative influence of the GHNs we classify the organizations according to their *official location* as declared in their Twitter account page. Usually, this location coincides with the headquarter of the organization. We are aware that the largest global organizations, such as the WHO, have branches in many countries and an official headquarter in a western country. In this study, for the sake of simplicity, we consider only the official location leaving to future developments the evaluation of the country level influence of each organization.

To understand the quality of the geographical coverage of the GHNs at the country level we need to compare the incidence of each disease as reported by official statistics with our data. As several countries appear to be mostly donors a simple yet effective way to distinguish is using the GDP per capita as a measure to discriminate donors and receivers.

In Figure 1 (top plot) we compare the structure of the international TBC network with the incidence of TBC by country in 2016. The map shows that African countries are most affected ones, followed by some Asian countries: Indonesia, Philippines and Mongolia, as well as the N.Korea dictatorship state. The size of the nodes is proportional to the number of organizations active in each country, whereas the links between organizations in different countries are reported in blue. Overall, the map shows a good coverage of source and destination countries. In the TBC network almost all regions of the world are well represented. This is due to the size of the network and to the large HIV/AIDS overlap. It is worth noticing, however, the fragility of some countries in south east Asia, such as Cambodia, Myanmar and Indonesia. Interestingly Mongolia has an high epidemics and still few official local organizations fighting the disease. North Korea have no organizations on Twitter for different reasons, while China lacks coverage on Twitter. The GPD per capita plot vs. number of NGOs of Fig. 1 confirms the fragility of poor countries such as Myanmar, Cambodia and Sierra Leone where the low income is associated with an insufficient number of NGOs.

Passing to the malaria network (Figure 2) we immediately see that some of the main

African countries by incidence of the disease are not covered. More in general, the network is smaller and there is a lack of coverage in some developed countries too, especially in Europe. We are aware that there can be bias in the Twitter data we use. However, the bias is likely to stay the same across disease areas. Therefore, by cross comparing Figure 1 and 2 we argue that the malaria GHN lacks of global coverage. We notice, in particular, the case of Tanzania where there is only one organization. Even though, we are not able to establish, from Twitter data, if the local presence is sufficient to fight the disease, we can detect a misalignment between the incidence of the disease and global coverage by Twitter. Finally the GDP per capita plot vs. the number of NGOs confirms the poor conditions of Sierra Leone but also Tanzania (where the incidence of Malaria is slightly lower). Countries such as Myanmar and Cambodia have no dedicated organizations for Malaria in our dataset.

In the pneumonia network (Figure 3) it is even more evident that the GHN does not have a global reach. Countries in western Europe, north America and western Africa are the backbones of the GHN, while nodes in south Asia appear to be almost disconnected. We interpret this as a sign of network weaknesses. The map depicts the percentage in death rates for children below 5 years old in 2016. Countries such as Philippines have high percentage of pneumonia death rates among children below five and just a single resident organization, the same is true for Bangladesh, a country with more than 160 millions inhabitants. China has a children mortality of 12 per cent that, given the size of the country, makes pneumonia - to the world level - a very serious disease. The plot of the GPD per capita vs the number of the NGOs describes the scarcity of organizations and also the small number of countries as donors. Our analysis confirms Pneumonia as a ”neglected disease”.

Figures 1,2,3 about here

5 Discussion and future work

During the recent years the increase in health aid has come mainly from global health networks of NGOs, international organizations, charities and global funds targeted to specific diseases and interventions. As a result GHN have attracted the attention of policy makers and health authorities as an important organizational solution to some of the main global healthcare challenges. However, the performances of GHNs have been different across diseases such as maternal mortality, neonatal mortality, pneumonia, TBC, malaria and so forth.

Currently, the role of the global networks as actors in international health governance is still questionable and the worldwide effects of the networks vary by issues. The search for new organizational arrangements, such as GHNs, is motivated by coordinate, organize and harmonize donor initiatives.

The main question is why some networks succeed in fighting a disease while others are far behind the expected outcome. It has been argued that the success or the failure of these global networks is strictly related to the fragmentation of global and national governance. However, so far there has been no comprehensive and transparent information about the network partnerships, internal procedures and often the objectives are long-term and difficult to measure.

The weaker networks show a poor governance and leadership, a scarce geographical coverage and a low level of connection. Conversely efficient networks mobilize large number of supporters, are able to influence the political and the cultural debate and more importantly they are able to produce changes in policies and in the re-allocation of resources.

Using SNA and machine learning techniques we reconstruct the GHN of malaria, TBC and pneumonia out of Twitter data. We analyze the role and position of each organization in the GHNs, their network power and centrality. Meanwhile for each account we measured the Twitter popularity. All these measures give us a clear insight of performance, overall success in the media campaigns and network leadership.

Results show that the more mature global network of TBC (partly overlapped with the AIDS/HIV one) reveals a structure characterized of high power through its size (1,838 nodes), connection and geographical coverage. The most influential nodes are the ones fighting for specific causes, for instance the international Aids vaccine initiative (IAVI).

Official organizations (such as the WHO) are present but they are not the most relevant in the network. The network is able to maintain an active structure, strong relationships correlated to an efficient service. Moreover it establishes a global coverage in all continents and countries.

The malaria network has less nodes (196 nodes) with respect to the TBC network. It is the most focused of the three networks: many nodes are acting as *propaganda machines* on Twitter, attracting a lot of attention. Initiatives such as RollBackMalaria, MalariaVaccine coordinate large groups of nodes in an effort to reduce the overall disease mortality. Their media campaigns on Twitter proved to be widespread and influential by number of people reached.

The Pneumonia GHN is the smallest one (65 nodes). It is a weakly and poorly connected network, lacking of geographical coverage and with a scarce level of focus on Twitter. Since only two nodes have an username related to the word pneumonia, it turns out that the most influential nodes of the network are funds, charities and official organizations (WHO) with no specific institution fighting pneumonia.

In this context the social network analysis on the Twitter platform may be of interest for the aid initiatives, poverty reduction and development strategies. For example the social networks have proved to be a invaluable instrument to engage stakeholders involved in health and raise funding for donors and health organizations. Among others Twitter emerged as the most open and free social network, researchers that can harvest the data about users and organizations with no strict privacy concerns and use these data to create measures of popularity and of efficacy for each media campaigns. Leaders in Twitter are those nodes being able to attract attention (more tweets) on the social media. The geographical information of the accounts allows to build maps of coverage for each initiative and to estimate the impact of each action. From this point of view extending our work to consider other - local - languages opens a large number of potential new insights and, for policy interventions, this make possible to plan better media campaigns.

Furthermore, every time local governments and international organizations are slow, social networks of NGOs can react quickly by using social media like Twitter to provide more timely and targeted help. This is specially true during emergencies where Twitter and the Internet become a reliable system for aid and relief coordination.

With a good media campaign gathering new users and attracting more donors is feasible to invest in special treatment regimes or in drugs which may be not available locally. Presenting the results of an action can further engage users, especially if they foresee a potential solution to the problem by supporting a specific aid program. From this perspective the social media are sophisticated health planning tools and they represent an important opportunity to increase both the quality and the quantity of aid for health at country and global levels.

Working with social media will improve the predictability of aid for health by addressing in the short and long term the needs of countries with limited donor support to increase the development assistance.

Finally, in poor countries the diffusion of the phone network and the basic level access to the internet can reward technologies such as Twitter In this case the Twitter social network can sustain local health plans to eradicate diseases. For instance simple methods to fight mosquitoes are of paramount importance to stop or slow down the diffusion of the malaria disease. Instructions on pesticides or nets can be diffused via social networks. Here again specific initiatives run by organizations can target Twitter on the local community level.

These tasks are challenging due to the size of the social media platforms, Facebook has more than 2 billions of users, and Twitter more than 300 millions: finding users within specific locations and their social role require a vast technological effort. Tools from machine learning, and artificial intelligence as well as a big data approach are needed to get lists of the most active users (the *influencers*), to measure popularity and attention level on the social media.

Future researches must improve the selection of users on Twitter considering not only the end users, or the organizations but also special categories such as practitioners (doctors, surgeons, health operators, investors etc.) and evaluate the total potential of the global health networks in terms of human workforce. Finally, we believe that the supply chain theory can be also applied to global health networks and aid networks. Models of network formation, total economic value, growth and development forecasts used in the supply chain domain explain why several companies are central in the production phase, and how each company can benefit from its position in the network.

If we are able to understand how the networks of organizations are forming in the early days, which kind of organizations and where (country or region) is more likely the formation of the leadership we can shape a better structure for GHN and aid. The goal of every action is to improve the aid effectiveness in health. The organizational challenge has to boost the development cooperation through an effective collaboration.

From a more theoretical perspective economic models used in the supply chain can be applied, with minimum modification, to understand success and failure of the GHNs. The challenge we addressed in this paper is network discovery and node role estimation. Creating a list of the actors involved in the aid/health network, defining their role and their intervention can be the starting point of future relevant analysis.

Finally our methodological approach, can be replicated to analyze other international social media networks targeting development goals such as nutrition, poverty, education and environmental protection.

6 Tables and figures

Table 1: Global network measures

disease	network density	num of edges	size
malaria	1.7%	641	196
pneumonia	7.3%	303	65
TBC	0.2%	7441	1838

Table 2: Average measures of the networks

disease	avg betweeness	avg closeness	avg clustering	avg degree
malaria	0.000166	0.025578	0.125212	6.540816
pneumonia	0.006483	0.147695	0.553756	9.323077
TBC	0.000089	0.023159	0.148521	8.096844

Table 3: Focus and overlap of the networks. The table shows for the disease networks, malaria, TBC and pneumonia the percentage of nodes having their username containing the work ”AIDS/HIV”, or ”pneumonia” or ”malaria”. As we considered the TBC network closely related with the HIV one, the total focus of the TBC network must read on the ”HIV/AIDS + TBC” column.

	HIV/AIDS	malaria	pneumonia	TBC	HIV/AIDS+TBC
malaria	3.6%	11.7%	0.0%	0.5%	4.1%
pneumonia	0.0%	3.1%	3.1%	3.1%	3.1%
TBC	4.5%	0.9%	0.1%	1.5%	6.0%

Table 4: Twitter popularity of the GHN networks

	avg tweets user	tot tweets	avg followers user	tot followers
netpneumonia	6906.0	448893	94445.9	6138982
netmalaria	4399.1	862225	57268.1	11224548
netTBC	2634.9	4842915	13124.3	24122476

Table 5: Network Leadership measures for the GHN

	organization	degree	closeness	betweeness
TBC	WorldVision	237	-	-
TBC	GMHC	233	-	-
TBC	PEPFAR	228	-	-
TBC	RedRibonFund	-	0.197	-
TBC	TBAlliance	-	0.195	-
TBC	stfrancishcs	-	0.174	-
TBC	WorldVision	-	-	0.0190
TBC	PEPFAR	-	-	0.0121
TBC	IAVI	-	-	0.0120
malaria	RollBackMalaria	72	-	-
malaria	GlobalFund	60	-	-
malaria	MalariaVaccine	55	-	-
malaria	RollBackMalaria	-	0.408	-
malaria	GlobalFund	-	0.320	-
malaria	bbbrieger	-	0.263	-
malaria	MalariaVaccine	-	-	0.011
malaria	EndMalariaNow	-	-	0.009
malaria	FightingMalaria	-	-	0.008
pneumonia	gatesfoundation	33	-	-
pneumonia	MSHHealthImpact	31	-	-
pneumonia	wellbeingafrica	29	-	-
pneumonia	GaviSeth	-	0.542	-
pneumonia	gatesfoundation	-	0.440	-
pneumonia	PeterASinger	-	0.413	-
pneumonia	WHO	-	-	0.133
pneumonia	MSHHealthImpact	-	-	0.074
pneumonia	wellbeingafrica	-	-	0.033

Table 6: Most popular accounts on Twitter for the GHN

	organization	followers-friends ratio	followers count	friends count
malaria	melindagates	5039	1048115	208
malaria	CDCemergency	3971	1827034	460
malaria	CDCgov	2885	784756	272
malaria	WHO	2746	3487656	1270
malaria	ClintonFdn	941	742014	788
pneumonia	WHO	2794	3585797	1283
pneumonia	gatesfoundation	1938	1808165	933
pneumonia	GaviSeth	155	28039	181
pneumonia	gatespoverty	93	22758	244
pneumonia	GlobalFund	45	165780	3692
TBC	KalamCenter	25373	1674661	66
TBC	melindagates	5468	1148227	210
TBC	drharshvardhan	4950	1499944	303
TBC	CDCemergency	3941	1832374	465
TBC	WildCare	3729	354231	95

Figure 1: Top: the global TBC network and TBC incidence 2016, new cases by country per 100000 inhabitants (WHO, 2016).The network is aggregated per country where a block model is superimposed. The size of the node is proportional to the number of organizations in that country. Bottom: the relationship between GPD per capita and number of NGO for every country: the lighter color of the nodes are associated to a higher incidence of the disease.

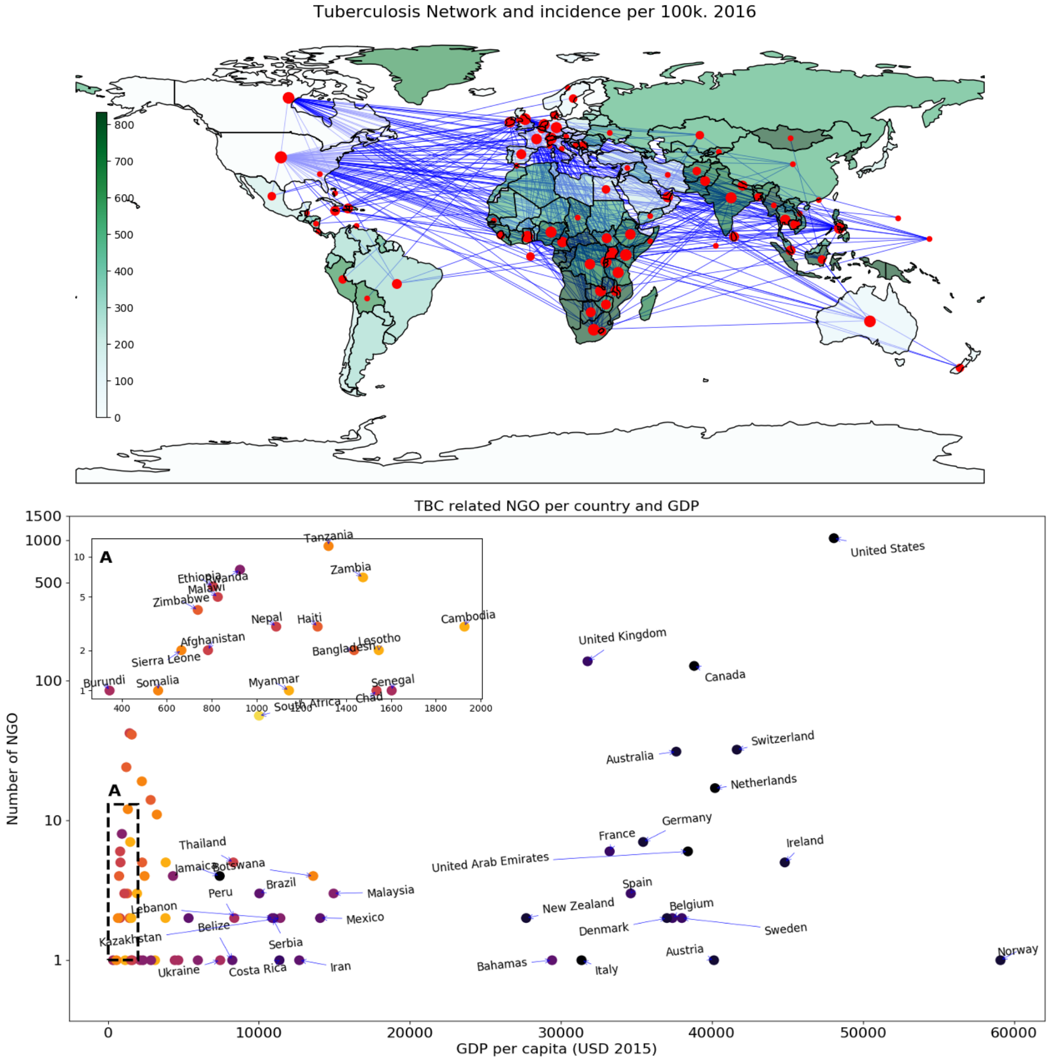


Figure 2: Top: the global malaria network and estimated malaria death by country per 100000 inhabitants (Global Malaria Mapper, 2017).The network is aggregated per country where a block model is superimposed. The size of the node is proportional to the number of organizations in that country.Bottom: the relationship between GPD per capita and number of NGO for every country: the lighter color of the nodes are associated to a higher incidence of the disease.

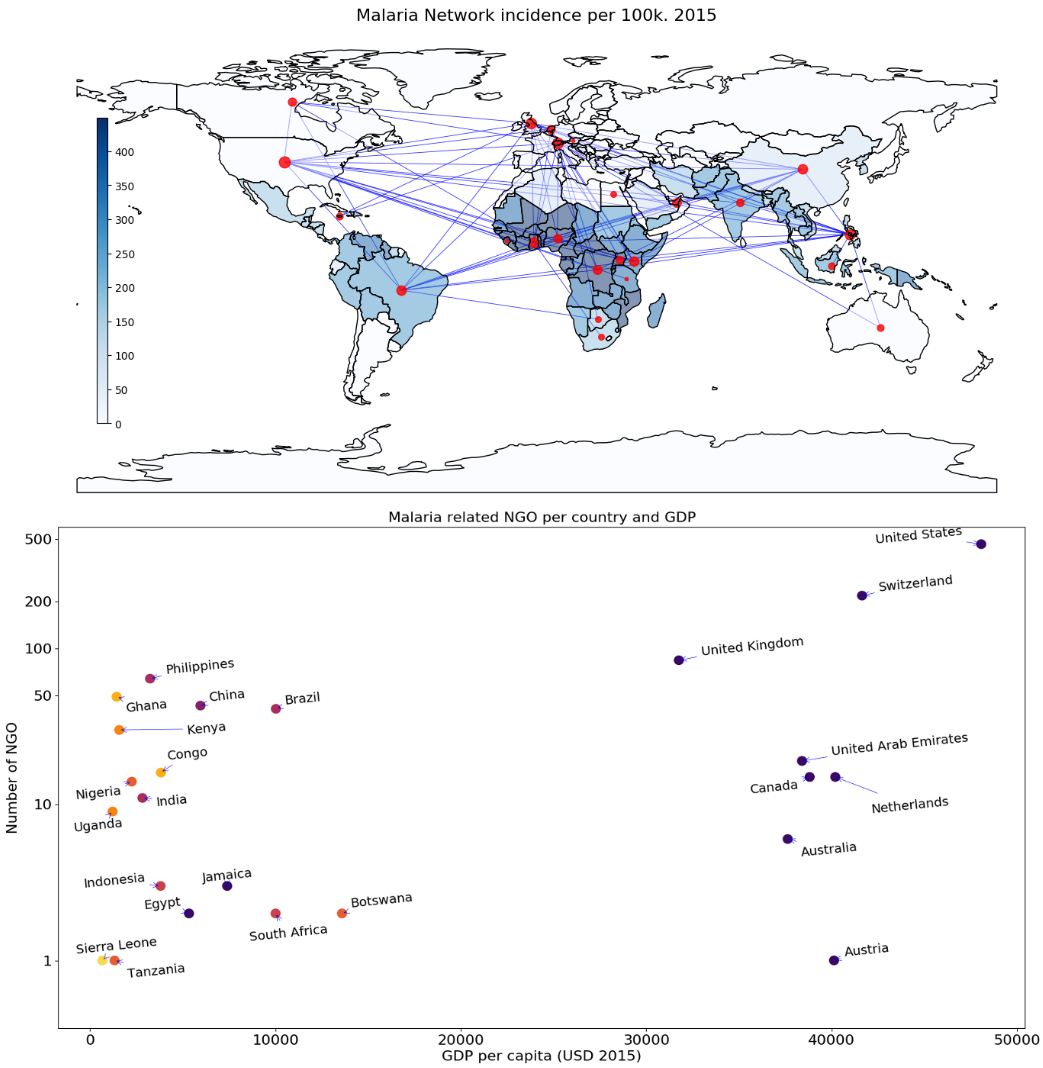
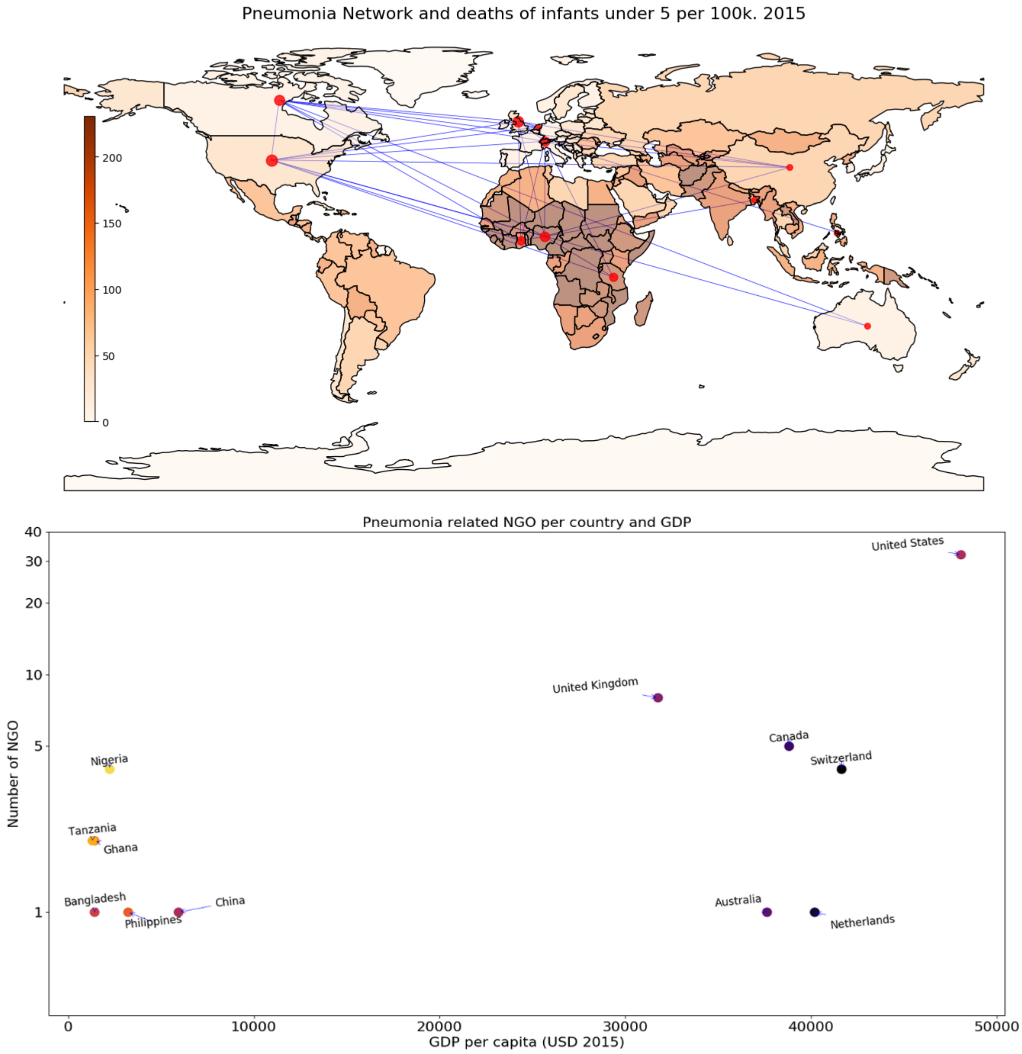


Figure 3: Top: the global pneumonia network and the percentage of deaths of children under five (UNICEF, 2017). The network is aggregated per country where a block model is superimposed. The size of the node is proportional to the number of organizations in that country. Bottom: the relationship between GPD per capita and number of NGO for every country: the lighter color of the nodes are associated to a higher incidence of the disease.



References

Acharya, A., A.T. Fuzzo de Lima, and M. Moore (2006). "Proliferation and fragmentation: Transaction costs and the value of aid." *The Journal of Development Studies* 42 (1): 1–21.

Annen, K. and S. Kosempel (2009). "Foreign aid, donor fragmentation, and economic growth." *The B.E. Journal of Macroeconomics* 9 (1): 1–30.

Bauer, Eric and Ron Kohavi (1999). "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants". *Machine Learning* 36 (1): 105–139.

Berlan, D. (2016). "Pneumonia's second wind? A case study of the global health network for childhood pneumonia." *Health Policy and Planning* 31 (1): 33–47.

Bokányi E., Lábszki Z. Vattay G. (2017). "Prediction of employment and unemployment rates from Twitter daily rhythm s in the US". *EPJ Data Science* 6 (1): 14.

Bollen, J., H. Mao, and X. Zeng (2011). "Twitter mood predicts the stock market". *Journal of Computational Science*.

Bueno, E., PM. Salmador, and O. Rodriguez (2004). "The Role of Social Capital in Today's Economy:Empirical evidence and proposal of a new model of intellectual capital". *Journal of Intellectual Capital* 5 (4): 556–574.

Burt, R. (1992). "Structural Holes: The Social Structure of Competition": 1–313.

Burt, R. S. (2005). "Brokerage & Closure": 1–296.

Chong, DPL. (2002). "UNTAC in Cambodia: A New Model for Humanitarian Aid in Failed States?." *Development and Change* 33 (5): 957–978.

Dershem L., T. Dagargulia L. Saganelidze S. Roels (2011). "NGO Network Analysis Handbook: how to measure and map linkages between NGOs". *Save the Children,Tbilisi, Georgia*.

Emirbayer, M. and J. Goodwin (1994). "Network Analysis, Culture and the Problem of Agency". *American Journal of Sociology* 99 (6): 1411–1454.

Emmerij, L. (2014). "The Future of Development Cooperation in Times of Crisis". *Development and Change* 45 (2): 384–394.

Ferguson, EW. et al. (1995). "A global health disaster network is needed". *BMJ: British Medical Journal* 310 (6991): 1412.

Firestone, R. et al. (2017). "the effectiveness of social marketing in global health: a systematic review". *Health Policy and Planning* 32: 110–124.

Fogel, RW. (2004). "Health, Nutrition, and Economic Growth". *Economic Development and Cultural Change* 52 (3): 643–658.

Freeman, LC. (1979). "Centrality in social networks conceptual clarification". *Social Networks* 1 (3): 215–239.

Gehring, K. et al. (2017). "Aid Fragmentation and Effectiveness: What Do We Really Know?." *World Development* 99 (11): 320–334.

Granovetter, MS. (1973). "The strength of weak ties". *American Journal of Sociology* 78 (6): 1360–80.

Gurajala, S. et al. (2016). "Profile characteristics of fake Twitter accounts". *Big Data & Society* 3 (2): 2053951716674236.

Hafner-Burton, EM., M. Kahler, and AH. Montgomery (2009). "Network analysis for international relations". *International Organization* 63 (3): 559–92.

Hanneman, RA. and M. Riddle (2005). “Introduction to Social Network Methods”. *University of California online pubblication*.

Heclo, H. (1978). “Issue networks and the executive establishment”. *in King A (ed). The New American Political System. Washington DC*.

Hong, L. and SE. Page (2004). “Groups of diverse problem solvers can outperform groups of high-ability problem solvers”. *Proceedings of the National Academy of Sciences of the United States of America* 101 (46): 16385–9.

Isett, KR. et al. (2011). “Networks in public administration scholarship: understanding where we are and where we need to go”. *Journal of Public Administration Research and Theory* 21 (1): 157–73.

Jablin, FM. and L. Putman (2001). *The New Handbook of Organizational Communication : Advances in Theory, Research and Methods*, 1–800.

Jones, B. (2017). “Looking Good: Mediatisation and International NGOs”. *The European Journal of Development Research* 29 (1): 176–191.

Kahler, M. (2009). “Networked politics: agency, power, and governance”. *Kahler M (ed). Networked Politics: Agency, Power, and Governance* 21 (1): 1–20.

Kimura H., Y. Mori and Y. Sawada (2012). “Aid proliferation and economic growth: A cross-country analysis.” *World Development* 40 (1): 1–10.

Kragelund, P. (2011). “Back to BASICs? The Rejuvenation of Non-traditional Donors’ Development Cooperation with Africa”. *Development and Change* 42 (2): 585–607.

Kulshrestha, Juhi et al. (2012). “Geographic Dissection of the Twitter Network.” *ICWSM*.

LaPorte, R. (1994). “Standards for information systems in the global health network”. *The Lancet* 344 (8937): 1640–1641.

Lecy, JD., IA. Mergel, and HP. Schmitz (2014). “Networks in public administration: current scholarship in review”. *Public Management Review* 16 (5): 643–55.

McEwan, L. and E. Mawdsley (2012). “Trilateral Development Cooperation: Power and Politics in Emerging Aid Relationships”. *Development and Change* 43 (6): 1185–1209.

Mocanu, Delia et al. (2013). “The twitter of babel: Mapping world languages through microblogging platforms”. *PloS one* 8 (4): e61981.

Munro, LT. (2005). “TFocus-Pocus? Thinking Critically about Whether Aid Organizations Should Do Fewer Things in Fewer Countries”. *Development and Change* 36 (3): 425–447.

Page, SE. (2007). “The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies”: 1–456.

Parkhe, A., S. Wasserman, and AD. Ralston (2006). “New Frontiers in Network Development”. *Academy of Management Review* 31 (3): 560–568.

Paxton, P. (1999). “Is Social Capital Declining in the United States? A Multiple Indicator Assessment”. *American Journal of Sociology* 105 (1): 88–127.

Podolny, JM. and KL. Page (1998). “Network forms of organization”. *Annual Review of Sociology* 24: 57–76.

Powell, WW. (1990). “Neither market nor hierarchy: networks forms of organization”. *Research in Organizational Behavior* 12: 295–336.

Pronk, JP. (2003). “Aid as a Catalyst: A Rejoinder”. *Development and Change* 34 (3): 383–400.

Quissell, K. and G. Walt (2016). “The challenge of sustaining effectiveness over time: the case of the global network to stop tuberculosis”. *Health Policy and Planning* 31 (1): 17–32.

S., Brindha, K. Prabha, and S. Sukumaran (2016). “A survey on classification techniques for text mining”. *2016 3rd International Conference on Advanced Computing and Communication Systems (ICACCS)* 1: 1–5.

Sabatier, PA. (2007). “Theories of the Policy Process”. *in 2nd edn. Boulder*.

Shiffman, J. et al. (2016). “A framework on the emergence and effectiveness of global health networks”. *Health Policy and Planning* 31: 3–16.

Sikkink, K. (2009). “The power of networks in international politics”. *Kahler M (ed). Networked Politics: Agency, Power, and Governance*: 228–47.

Swamy, P., K. Ngo, and A. Mandalakas (2017). “Tick, Tock, Clock: When is the best time to post on Twitter?” *Annals of Global Health* 83 (1): 165–166.

Trevor Thrall A. D. Stecula, D. Sweet (2014). “May We Have Your Attention Please? Human-Rights NGOs and the Problem of Global Communication”. *The International Journal of Press/Politics* 19 (2): 135–159.

Weiss, CH. (1972). “Evaluation Research: Methods for Assessing Program Effectiveness”: 1–239.

White, H. and L.Woestman (1994). “The Quality of Aid: Measuring Trends in Donor Performance”. *Development and Change* 25 (3): 527–554.

White, H. and M. McGillivray (1995). “How Well is Aid Allocated? Descriptive Measures of Aid Allocation: A Survey of Methodology and Results”. *Development and Change* 26 (1): 163–183.

WHO (2007). “Aid effectiveness and health”: 1–25.

WHO, Report (2016). “Revised WHO classification and treatment of childhood pneumonia at health facilities”: 1–34.

Wholey, JS. (1983). “Evaluation and Effective Public Management”: 1–34.

Zhongming M. P. Gautam, ORL. Sheng (2011). “Mining Competitor Relationships from Online News: A Network-based Approach”. *Electron. Commer. Rec. Appl.* 10 (4): 418–427. ISSN: 1567-4223.

IMT SCHOOL FOR ADVANCED STUDIES LUCCA

Piazza San ponziano 6, 55100 Lucca, Italy.
www.imtlucca.it