scientific reports

OPEN

On topological properties of COVID-19: predicting and assessing pandemic risk with network statistics

Mike K. P. So[1✉], Amanda M. Y. Chu[2], Agnes Tiwari[3,4] & Jacky N. L. Chan[1]

The spread of coronavirus disease 2019 (COVID-19) has caused more than 80 million confirmed infected cases and more than 1.8 million people died as of 31 December 2020. While it is essential to quantify risk and characterize transmission dynamics in closed populations using Susceptible-Infection-Recovered modeling, the investigation of the effect from worldwide pandemic cannot be neglected. This study proposes a network analysis to assess global pandemic risk by linking 164 countries in pandemic networks, where links between countries were specified by the level of 'co-movement' of newly confirmed COVID-19 cases. More countries showing increase in the COVID-19 cases simultaneously will signal the pandemic prevalent over the world. The network density, clustering coefficients, and assortativity in the pandemic networks provide early warning signals of the pandemic in late February 2020. We propose a preparedness pandemic risk score for prediction and a severity risk score for pandemic control. The preparedness risk score contributed by countries in Asia is between 25% and 50% most of the time after February and America contributes around 40% in July 2020. The high preparedness risk contribution implies the importance of travel restrictions between those countries. The severity risk score of America and Europe contribute around 90% in December 2020, signifying that the control of COVID-19 is still worrying in America and Europe. We can keep track of the pandemic situation in each country using an online dashboard to update the pandemic risk scores and contributions.

Since the declaration of the coronavirus disease 2019 (COVID-19) as a global pandemic by the World Health Organization (WHO) on 11 March 2020, there have been more than 80 million confirmed cases and more than 1.8 million deaths worldwide as of 31 December 2020. Many countries have imposed preventive measures, such as city lockdown, travel restrictions, quarantine, widening social distancing, enhancing personal hygiene, etc, for stopping or slowing down further transmission of the disease. While continuing with the pandemic measures, it is also important to assess the pandemic risk for setting up preparedness plan and control measures for COVID-19[1] especially for those countries with high risk of wide-spread transmission[2].

A common pandemic risk assessment approach is by epidemiological modeling, where population is divided into at least three groups: susceptible, infected, and recovered[3–6]. This approach has been successful in understanding the disease transmission in a region through mathematical formulation of human interactions which may lead to infection, and the rate of recovery from infection. However, the susceptible-infection-recovery approach was designed fundamentally for closed populations[7] and so restricted to regional studies. To supplement the epidemiological modeling, it is helpful to combine data from different regions or countries to study the pandemic situation, i.e. the status of the disease transmission in multiple regions or over the world, and to quantify the potential risk involved due to the COVID-19 outbreak. Two issues are of particular interest: prediction and control. Prediction means whether we can learn insights from data to produce early warning signals of pandemic and for preparedness[8–10]. Control refers to assessing the current pandemic risk severity for deciding proper measures during global pandemic[11,12].

In this paper, we propose network analysis[13] to aggregate data information from all over the world to construct dynamic pandemic networks for COVID-19. Network analysis has been applied to analyzing scientific

[1]Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology, Hong Kong, China. [2]Department of Social Sciences, The Education University of Hong Kong, Hong Kong, China. [3]LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [4]School of Nursing, Hong Kong Sanatorium & Hospital, Hong Kong, China. ✉email: immkpso@ust.hk
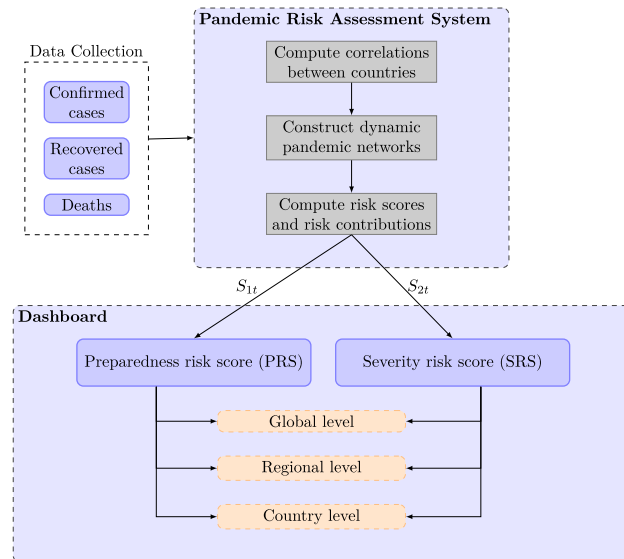
Figure 1. The pandemic risk assessment system work flow diagram.

collaborations[14], to econometrics for assessing systemic risk and contagion effects in financial markets[15,16], to medical research in studying gene co-expression, disease co-occurrence and global epidemics[17–20], and to analyzing text[21]. We extend the network approach in the literature[9,22] to study topological properties of COVID-19. There are three main features of our proposed network analysis. First, we make use of publicly available data, namely daily number of confirmed cases and daily accumulated number of infected people in each country to learn topological properties of dynamic pandemic networks and to visualize the propagation of COVID-19 for risk prediction and control. Second, we construct two pandemic risk scores and determine risk contributions from countries for pandemic prediction and control. The risk contributions of countries are helpful for setting preparedness plans and for assessing the severity of COVID-19 outbreak in specific regions. Finally, we compare the COVID-19 topological features with an independent network, the Erdos–Renyi model[23], to understand the current pandemic situation and observe any signal for the COVID-19 outbreak to go away (for restarting economic activities).

In the healthcare management consideration, it is important, yet challenging, to assess pandemic risk of COVID-19 in the global perspective. In this study, we propose the network analysis and two pandemic risk scores using publicly available information on the number of confirmed COVID-19 cases, and the estimated number of currently infected people. We provide evidence that network statistics can yield early warning signals of the outbreak of COVID-19. The time series of the two pandemic risk scores and the respective risk contributions by countries can help to predict or 'stress test' the potential risk involved from releasing travel restriction measures between countries, and to estimate the severity of the pandemic for epidemic control. Simulations from a network assuming independent links between countries suggest that there is no sign for COVID-19 to go away in a few months. In summary, Fig. 1 shows the work flow diagram of the pandemic risk assessment system developed in this paper. We started from publicly available confirmed cases, recovered cases and the number of deaths due to COVID-19, to constructing dynamic pandemic networks and their network statistics, and finally compiling the preparedness risk score (PRS), and the severity risk score (SRS) for risk assessment. Details of the work flow are given in Fig. 1 and in "Methods". For tracing the current pandemic risk of 164 countries, we have included the most updated risk scores and risk contributions in http://covid-19-dev.github.io/.

Results

Network connectedness. As described in "Methods", we followed the literature[9,22] to construct time series of pandemic networks of 164 countries from February 2020 to December 2020. The countries were linked together in a particular day when the correlation of the increase in the number of confirmed cases in the past 14 days exceeds a certain level. This 'co-movement' of newly confirmed COVID-19 cases defines the links between countries and forms the pandemic networks. To quantify the network connectedness, we calculated the network statistics presented in "Methods". Figure 2 shows the pandemic networks on 4 February 2020, 11 March 2020, and 11 April 2020. From the figure, we can visualize the pandemic networks are getting more dense from February to April 2020. This visual increase in connectedness was documented as an early warning signal of the COVID-19 pandemic in the literature[22]. The left panel of Fig. 3 presents the number of edges, network density, clustering coefficient, and the assortativity coefficient of the COVID-19 pandemic networks from February 2020 to December 2020. The right panel of Fig. 3 gives the corresponding network statistics for the independent network, the Erdos-Renyi model, which are discussed in the *Discussion*. From the left panel of the figure, the actual number of edges in the pandemic network increases steadily from late February 2020 and rises sharply after the WHO's declaration of the COVID-19 pandemic on 11 March 2020 to the peak on 18 March 2020. The sharp
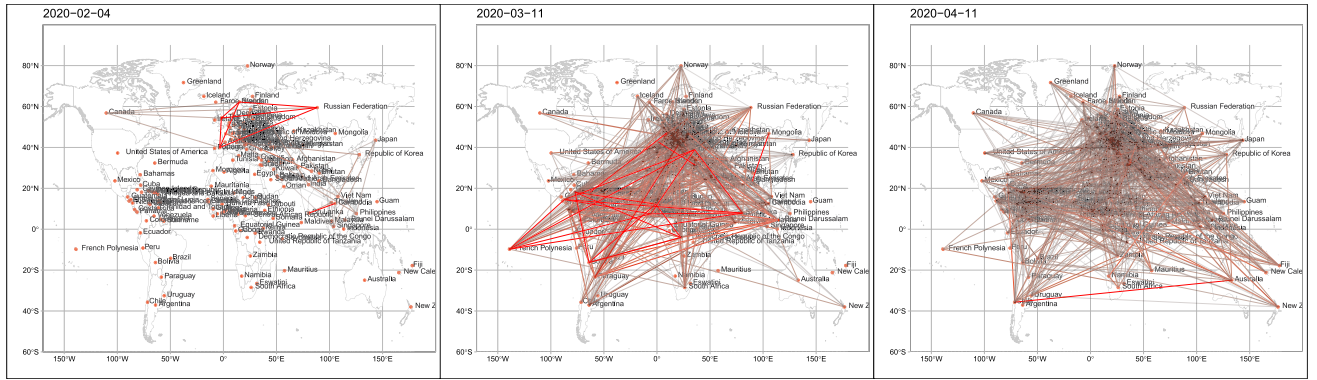
Figure 2. The COVID-19 pandemic networks on 4 February 2020, 11 March 2020, and 11 April 2020.
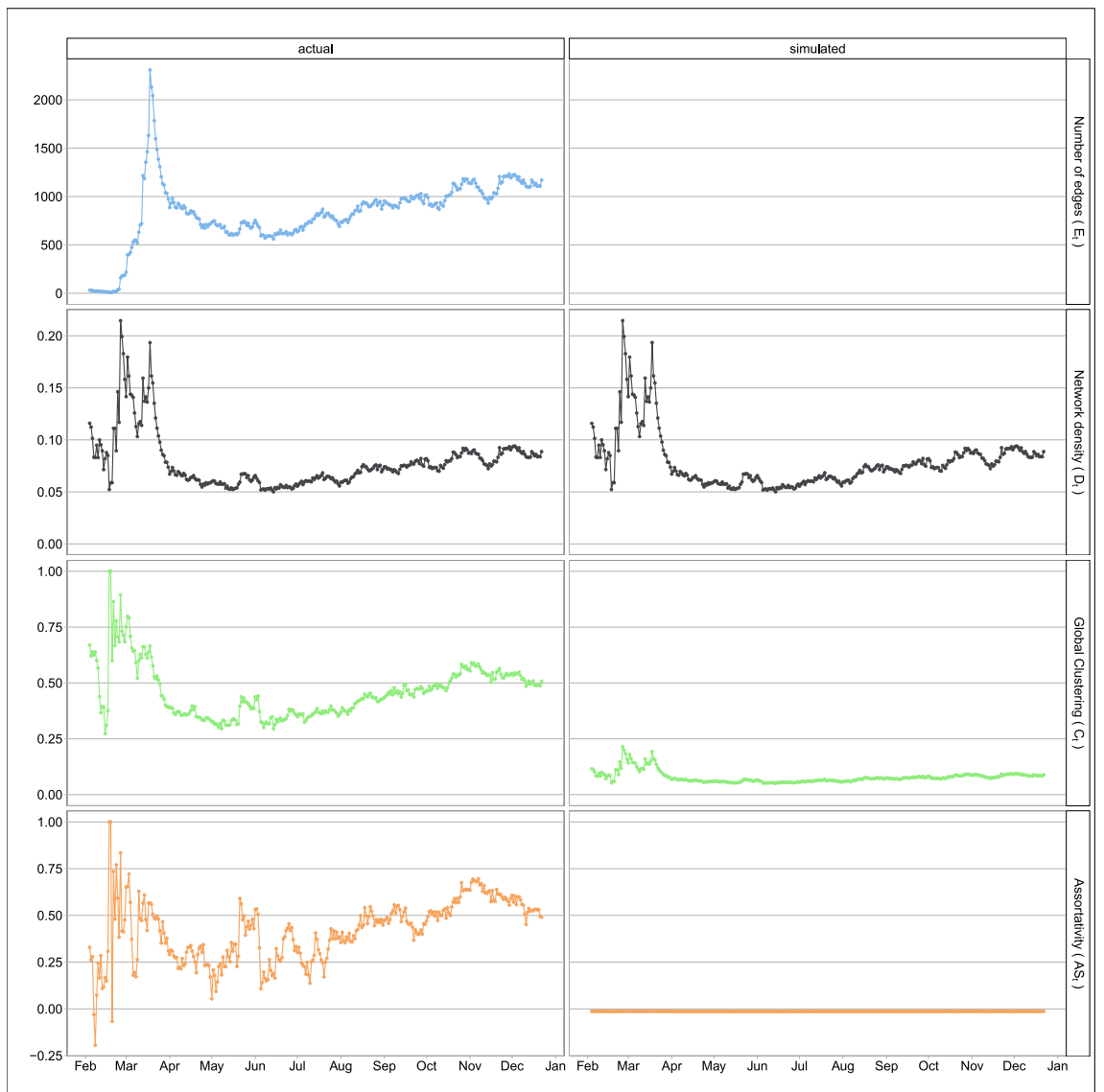

Figure 3. The time series plots of the network statistics, including the number of edges, network density, clustering coefficient and assortativity coefficient. The left column gives the network statistics for the COVID-19 pandemic networks, and the right column gives the corresponding statistics from an independent network, the Erdos–Renyi model (simulated networks).

increase of the number of edges in the week of 12–18 March 2020 is partly attributed to having more countries reporting their confirmed COVID-19 cases, and partly due to more edges being formed because of the 'common trend' in the confirmed cases in more countries. The pattern of the number of edges in this particular week evidently confirms the pandemic announced by the WHO. The downward trend in the number of edges after 18 March 2020 signifies the time period of epidemic control of the further spreading for COVID-19 by countries which implemented various measures, like quarantine rules, travel restrictions, and enlarging social distancing. The downward trend also reveals the effect of those measures on the pandemic control and thus seeing less coherence in the occurrence of confirmed cases in the countries. The number of edges has been stable in April to June 2020, but there is a rising pattern in July to December 2020.

The research in the literature[9] proposed that we can use the network density, D_t in Fig. 3, to generate early warning signals of the COVID-19 pandemic. We observe in the left panel of Fig. 3 that D_t shows the first peak on 26 or 27 February 2020 which is analogous to 'pre-earthquake phenomena' as an alarm to a big earthquake some days later. The first peak of the network density, induced by the high connectedness of changes in confirmed COVID-19 cases from countries in late February, probably tells us that the pandemic was happening. There has been stronger evidence of co-occurrence of more confirmed cases in the countries, leading to a sharp increase in the network density in just two to three days in late February. There is another peak in the network density in mid March after the WHO's declaration of the COVID-19 pandemic on 11 March 2020. Similar to the number of edges, the network density is quite stable in April to June and has an increasing trend in July to December 2020.

The clustering coefficient at time t is a measure of how close the pandemic network at time t to a perfectly-linked network where all countries are linked together. In the extreme case (of the perfectly-linked network) where all countries' confirmed cases increase simultaneously in the past 14 days, the countries will be well connected and can be viewed as one big pandemic region. Therefore, when the clustering coefficient is higher, the pandemic network will be more similar to the perfectly-linked situation, indicating stronger evidence of the global pandemic. From the clustering time series in the third row of the left panel in Fig. 3, we observe that the highest value occurs on 26 February 2020, when we also observe the first peak in the network density. The clustering coefficients also present an early warning signal of the global COVID-19 pandemic. Unlike the network density, the clustering coefficients do not show another obvious peak in mid March. Although we still observe substantial confirmed COVID-19 cases in various countries, the clustering coefficients on around 18 March (when we see another peak in the network density) is smaller than that on 26 February 2020, implying that the pandemic network is less similar to the perfectly-linked network than what we see on 26 February 2020. The downward trend in the clustering coefficient probably implies that the global outbreak of COVID-19 was gradually controlled since early March, though it takes time for the actual confirmed cases to go down substantially. The clustering coefficient lies between 0.25 to 0.50 in April to September 2020 and exhibits an increasing trend after June 2020. It exceeds 0.5 for a few weeks in late October to early November 2020.

The assortativity coefficient, AS_t in Eq. (5), is to assess the assortative mixing of the vertices or countries in the pandemic networks. The last row in Fig. 3 (in the left panel) presents AS_t from February 2020 to December 2020, where we see the value of AS_t to jump up to around 0.5 in late February. The substantial increase in AS_t in February implies that countries with a large number of connections tend to link with high-connection countries. This pattern of association of 'similar' countries (in terms of the degree k_{it} in Eq. (3)) can indicate possible inflection risk in social networks[24]. In the current COVID-19 pandemic situation, we can view this high AS_t as an indication of the disease outbreak since it is likely when a global pandemic occurs, high-connection countries tend to be linked together because they may have a simultaneous increase in the number of confirmed COVID-19 cases in a short period of time. It is not surprising to see that the time series pattern of AS_t is similar to that of C_t in the third row of Fig. 3. The assortativity coefficient exhibits a decreasing trend from early March to early May 2020, and lies between 0.25 and 0.5 most of the time in May to September 2020. It stays at above 0.5 in November and early December 2020, indicating quite high assortative mixing in these two months.

Pandemic risk sores. The preparedness risk score (PRS), S_{1t} and the severity risk score (SRS), S_{2t} are computed for t from 4 February 2020 to 29 December 2020. Both ω_t and n_t are measured in millions of people. We plot the standardized S_{1t}, in which S_{1t} is divided by the total number of possible interactions (when all countries are directly linked and ω_t taken as the population sizes) in Fig. 4 (left top panel). The first peak appears on 2 March 2020 when the standardized S_{1t} increases sharply from around zero to close to 0.1 in just a few days. Using 0.05 as a reference, i.e. 5% of the total interactions between people in the 164 countries, this sharp increase marks the first time when the risk score exceeds this reference, and can be regarded as an early warning signal of the COVID-19 pandemic. The second peak is found on 18 March 2020, a week after WHO's declaration of COVID-19 as a global pandemic on 11 March 2020. This spike reaches a point close to 0.4, accounting for more than one-third of the total interactions between people in the 164 countries. After that, probably due to stringent measures imposed by various countries, including travel restrictions, community lockdown and enhancing social distancing, the standardized S_{1t} drops but quite slowly in mid April and May. The time series of the PRS stays mostly above 0.05 till mid June, indicating that the pandemic risk is still substantial even after three months of the WHO's declaration with tremendous measures and efforts from various countries in preventing the transmission of COVID-19. An increasing trend appears in the standardized PRS in June to November 2020, when we see a local maximum of above 0.2. What worries us is whether this standardized S_{1t} will go up again after December 2020 if some countries release travel restriction measures later.

Figure 4 (left bottom panel) displays the time series of the standardized S_{2t}, in which S_{2t} is divided by the total number of possible interactions (which all countries are directly linked, with ω_t taken as the population sizes and n_t taken as 0.1% of the population sizes). The construction of S_{2t} is more sensitive to the confirmed COVID-19 cases and thus S_{2t} can serve as a measure of the COVID-19 outbreak severity. The first peak of S_{2t}, signifying
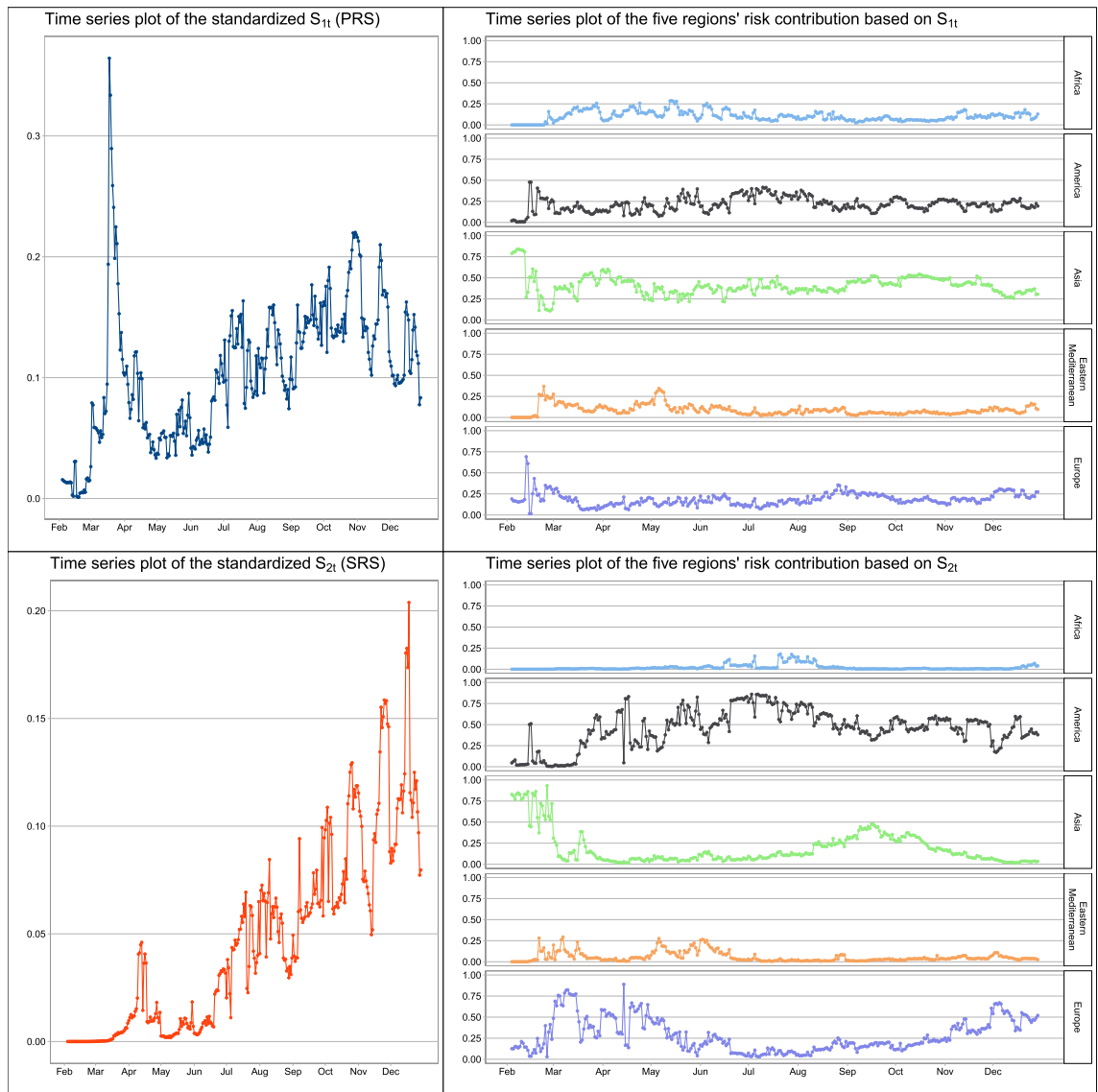
Figure 4. Plot of the standardized S_{1t} with the standardization done by all possible interactions between people in the 164 countries, and the standardized S_{2t} with the standardization done by all possible interactions between people in the 164 countries divided by 0.1%.

the first wave of the pandemic, is on 12–13 April 2020, roughly a month after the WHO's declaration on 11 March 2020. There is improvement in reducing the number of confirmed cases and possibly also the pandemic network connectedness that we observe obvious drop in S_{2t} in late April and early May. However, another local peak is observed on 30 May 2020, which likely indicates the second wave of the outbreak in the second half of May. The recent increasing trend of S_{2t} appeared in late June to December 2020 probably implies that the third wave of the outbreak of COVID-19 has come. The situation may get worse later if there is no preventive measure for further stopping the transmission of the disease in local community or if travel restrictions are going to be suspended or stopped.

Risk contributions from countries. Figure 4 (right top panel) gives the risk contributions of Africa, America, Asia, Eastern Mediterranean and Europe by grouping the f_{it}, the risk contribution based on the first pandemic risk score S_{1t} in Eq. (8), in the five regions. We observe that Asia contributes most of the pandemic risk associated with S_{1t} in February 2020 and the Asia's contribution stayed mostly between 0.25 and 0.5 after February. In the predictive perspective, reducing the mobility of the Asia's population can be a reasonable preventive measure during the pandemic. There is an upward pattern in the America's contribution to close to 0.5 in mid June to July 2020 and in the Europe's contribution to above 0.25 in late August and December 2020, indicating relatively higher risk contribution of transmission of the disease due to possible interaction between susceptible populations.

Figure 5 presents the heatmap of f_{it}. Based on the construction of S_{1t} using mainly the interactions between two susceptible populations, this risk contribution is more on preventive preparation if there is chance for people
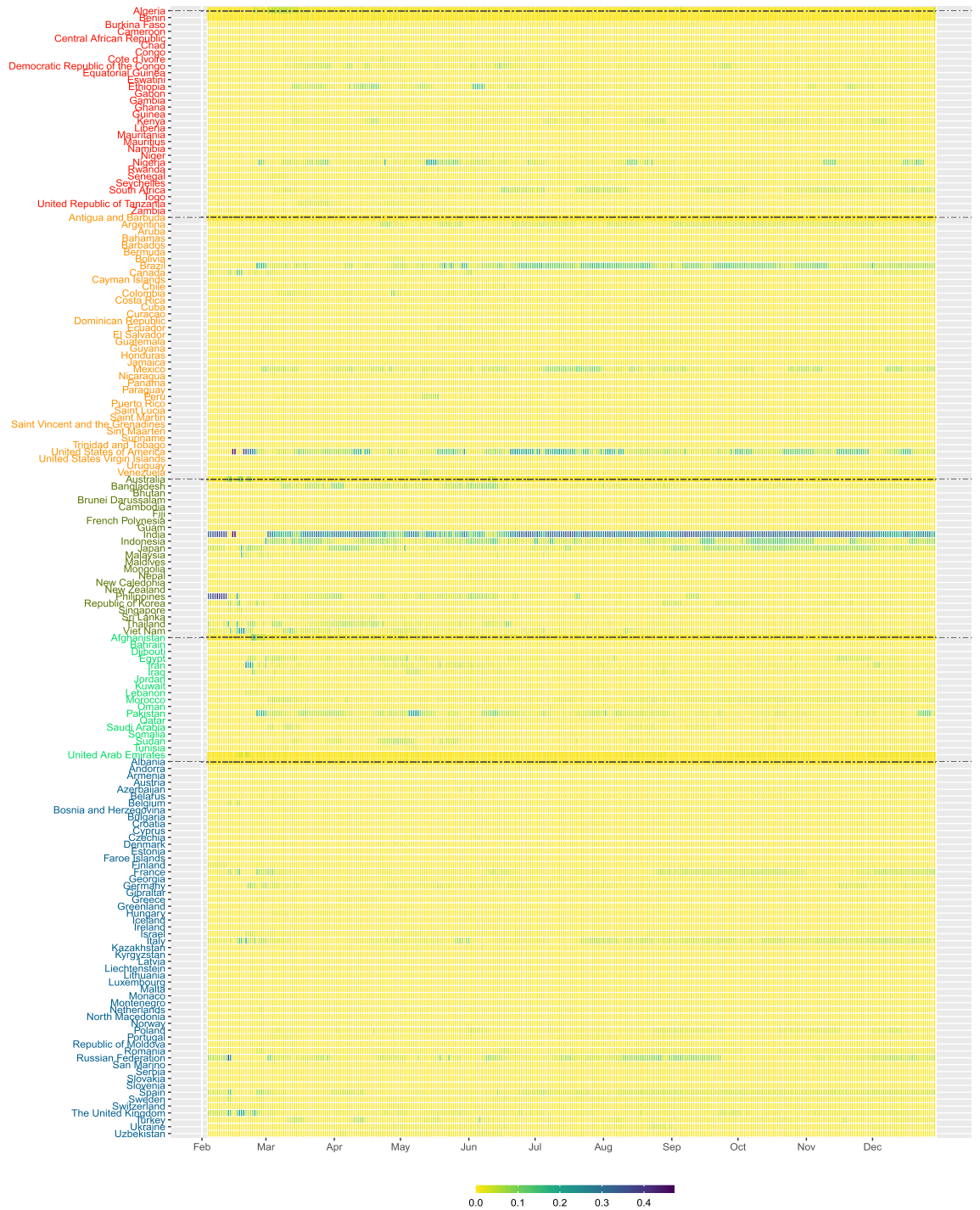
Figure 5. The heatmap of the risk contribution f_{it} based on the preparedness risk score S_{1t} for the 164 countries from February to December 2020.

from the two countries to interact. So the risk contribution f_{it} is like doing 'stress testing' of the pandemic risk. In early February, the risk score is mainly contributed by Asia, where the risk contributions from India and Philippines are around 0.3. From late February to early March, the risk score from non-Asian region, e.g. United States from America, Iran and Pakistan from Eastern Mediterranean, and Italy, Germany and France from Europe started showing significant contributions to the risk score. In particular, f_{it} of the United States has been quite high since mid February, signifying the early stage of the pandemic in the country. In late May and June, Ethiopia, Nigera from Africa, and Brazil from America also showed around 0.1 contribution that should be alerted. The risk contribution f_{it} of the India, Indonesia, Japan in Asia, Brazil and the United States in America, and France, Italy and Russia in Europe are quite relatively high in July to November 2020.
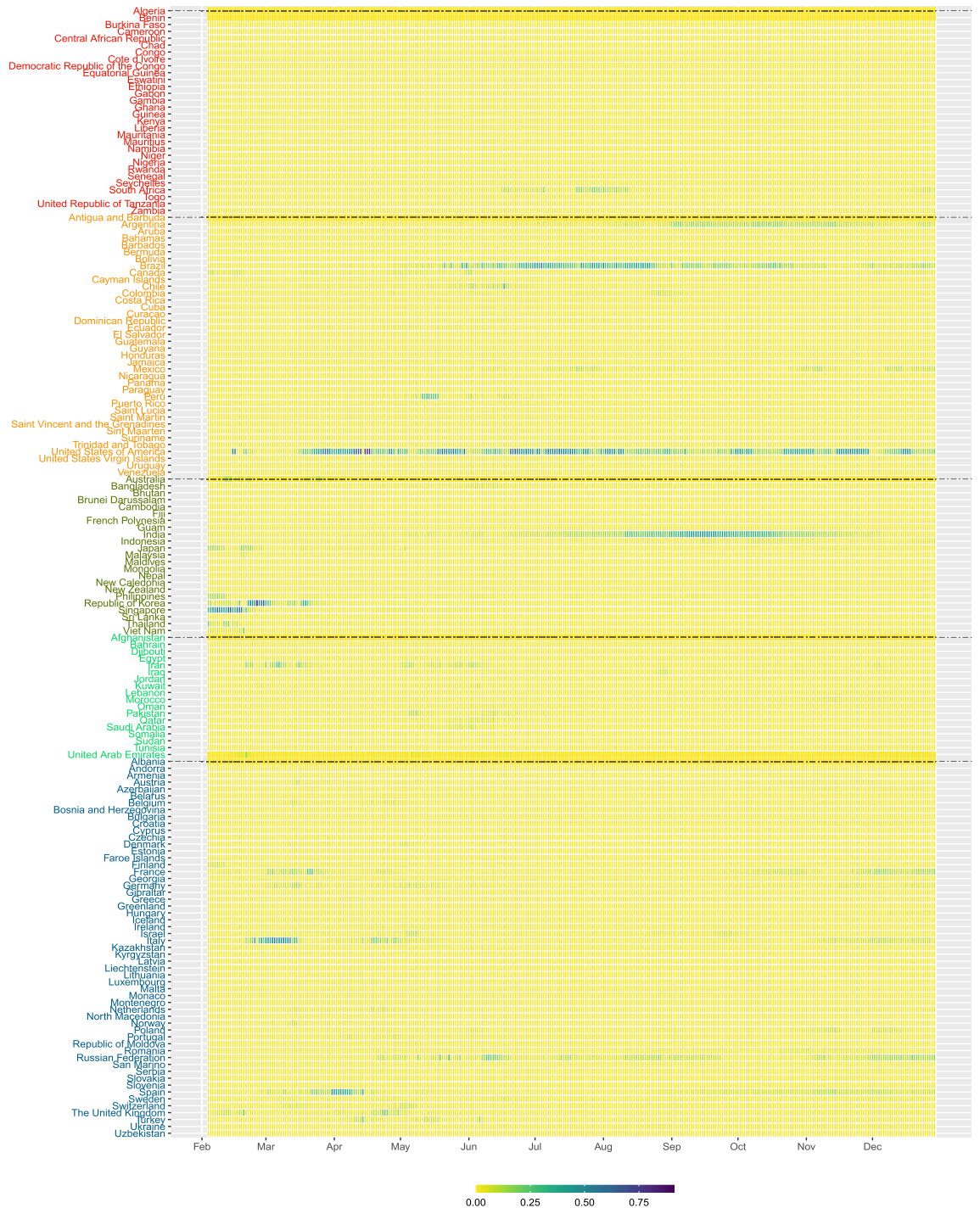
Figure 6. The heatmap of the risk contribution g_{it} based on the severity risk score S_{2t} for the 164 countries from February to December 2020.

Figure 4 (right bottom pannel) gives the risk contributions of Africa, America, Asia, Eastern Mediterranean and Europe by grouping the g_{it}, the risk contribution based on the second pandemic risk score S_{2t} in Eq. (9). Since S_{2t} is deduced from the number of possible interactions between people of a susceptible population and people of an infected population, the number of confirmed cases will play an important role in determining the risk contribution g_{it}. From Fig. 4, the pandemic risk based on S_{2t} was contributed mostly by Asia in February 2020, and mostly by Europe and America in March and April 2020. After April, the risk contribution from America climbed up to about 0.5, and exceeded 0.75 in July 2020. In September and October 2020, the risk contribution is dominated by Asia. The risk contribution of America and Europe are at high levels in November and December 2020, when the winter comes. Figure 6 presents the heatmap of g_{it}. This risk contribution reflects the severity of the COVID-19 pandemic risk from each country and thus is a viable measure for guiding us on the pandemic

7

control of further transmission of the disease. In early February, the risk contributions of Japan, Philippines, Korea, Singapore and Thailand are quite high. The risk contributions are then dominated by Korea, Iran and Italy in early March, and followed by United States and Spain in late March. We can see that f_{it} of United States appears to be high in late February, earlier than seeing large value in g_{it}, probably indicating that there was chance for United States to set up stringent control measures in late February before the widespread of the disease reflected in g_{it} afterward. On the contrary, India decided the national lockdown policy on 24 March 2020[25]. The lockdown policy may account for the fact that even though its f_{it} showing quite high value in February, its risk contribution indicated by g_{it} is very low in March and April. The cases of United States and India probably indicates the possible use of f_{it} as a measure for setting up timely stringent measures. In April to June, the risk contribution g_{it} is dominated by Brazil, Peru and United States in America, and Italy, Russia and Spain in Europe. India and South Africa's risk contributions have increased in July and August, respectively. The risk contribution of European countries are quite high in December 2020. The severity of the COVID-19 pandemic risk in those countries with relatively high g_{it} cannot be ignored.

Discussion

To assess the current pandemic status of COVID-19, we simulate networks from the Erdos-Renyi model[23]. Specifically, at each time t, we generate 10,000 networks with edges independent of each other. To maintain the same level of network density, two vertices in a simulated network are linked together with the probability D_t. The right panel of Fig. 3 displays the time series of the average of the network statistics from the 10,000 simulated networks. The outbreak of a disease likely makes the pandemic network to follow some structures which deviate from the Erdos-Renyi model. For example, in the COVID-19 pandemic, the common trend in the confirmed COVID-19 cases observed between countries i and j and countries j and k may imply a relatively higher chance to observe a common trend on the COVID-19 confirmed cases in countries i and k as well (this is the transitivity property in social networks). Therefore, by comparing the discrepancy between the time series patterns of C_t and AS_t between the COVID-19 pandemic networks and the Erdos–Renyi model, we can find hints as to when the global COVID-19 pandemic will go away. From Fig. 3, the clustering coefficients of the pandemic networks are well above that of the Erdos–Renyi model with the same network density D_t. Although there is sign for the assortativity coefficients of the pandemic networks to approach zero, the assortativitiy coefficient of the Erdos–Renyi model, in June and July, it goes up again after July 2020 and reaches a local maximum in November 2020. Comparing D_t, C_t and A_t of the pandemic networks and the same network statistics simulated from the Erdos–Renyi model, there is no strong evidence that COVID-19 will go away shortly after December 2020.

The preparedness risk score S_{1t} accounts for the possible interactions between susceptible population of countries being linked together in the pandemic networks, and can be used to quantify the risk of transmission of the COVID-19 between countries. The possible transmission risk is attributed to substantial percentages of asymptomatic transmission or presymptomatic transmission[26]. Therefore, this S_{1t} score is a kind of 'stress testing' measure[27] which evaluates the potential transmission risk involved if travel restriction policies are canceled or become less strict. From the time series plot of the standardized S_{1t} in Fig. 4, the transmission risk, or S_{1t}, increases sharply on 2 March 2020 to 0.08, and 18 March 2020 to 0.36, implying more than 1/3 of all possible interaction between susceptible populations of the 164 countries. The first peak on 2 March 2020 (well before the WHO's declaration on 11 March 2020) can be regarded as an early warning signal of the COVID-19 global pandemic. The second peak on 18 March 2020 marks the rapid spreading of the disease in March and early April, when the number of cases of Italy, Spain, France and other European countries increase substantially[3,28,29]. After reaching the peak on 18 March 2020, the transmission risk went down in April and May, probably because of a series of travel restrictions and other lockdown measures which significantly stopped the mobility of susceptible population between countries to almost completely disappeared. However, S_{1t} didn't go down further and stayed at around 0.05 in April and May, implying that the transmission risk had not been reduced to a level that would enable the release of travel restriction measures. Even worse is that the increase in the transmission risk is reignited in late June and July and continues to be quite severe till December 2020.

The risk contribution f_{it} in Eq. (8) based on S_{1t} quantifies the relative impact of country i at time t in transmitting COVID-19 through interaction between susceptible populations. The higher the value of f_{it}, the higher is the potential impact from country i at time t with respect to the transmission risk. The risk contribution f_{it} can be interpreted as a relative transmission risk measure which can be considered by governments to set up preventive measures by revisiting their epidemic control policies applied to different regions from time to time. For example, we observe that f_{it} of Nigeria and South Africa (of Africa), Brazil, Mexico and United States (of America), India and Indonesia (of Asia), Pakistan (of Eastern Mediterranean) and Russia and Spain (of Europe) are relatively high in June and July. Governments are recommended to be more alerted when revisiting any preventive measure with those countries whose f_{it}'s are high.

The severity risk score S_{2t} calculates the possible interactions between susceptible populations and the currently infected population of countries being linked together in the pandemic networks. The score reflects the current severity of the COVID-19 outbreak due to the transmission of disease from infected people who may have symptoms or due to presymptomatic transmission. The evolution of pandemic risk severity reflected by the time series pattern of S_{2t} provides a reference for governments to revisit their current epidemic control measures to see if it is necessary to revise or to strengthen the current measures for preventing their medical systems from overloading too much or breaking down. In Fig. 4, we observe from S_{2t} the first wave of COVID-19 outbreak in early to mid April when there is substantial increase in the confirmed COVID-19 cases in Europe[28]. There is an obvious drop in S_{2t} in late April till mid May, probably due to the lockdown control in European countries[30]. The second wave of the COVID-19 outbreak fell in late May when there were graduate economic reactivation activities implemented. Another sharp increase in the risk score started in late June and the increasing trend continued

in July to December 2020, signifying the third wave of the outbreak. The high value in S_{2t} can give an alerting signal to governments about the need to slow down the recovering of social activities or the revival of economy.

Although g_{it} refers to the risk contribution due to interactions between infected people and susceptible population of countries, high g_{it} also implies high risk of 'domestic transmission' if people from different regions of a country can have physical contact, for example, through school resumption, social, or religious activities. Before and during the first wave of COVID-19 outbreak in March and April, we see relatively high g_{it} in Fig. 6 in France, Germany, Italy, Spain, United Kingdom and United States. After imposing strict countrywise lockdown policies, the risk contributions of many European countries, except Russia, went down in May. In terms of g_{it}, the pandemic risk severity of United States does not significantly lower in May. We also see the transmission of the disease to South America where the risk contribution of Peru is quite high in May. In July, the total risk contribution in Brazil and United States accounts for around 80% of the overall S_{2t} and we also see substantial increase in risk contribution in South Africa. The risk contribution of India is the highest in September and October 2020. The g_{it} heated up again in the United States and European countries in November and December 2020. Governments and healthcare specialists may not ignore relatively high g_{it} which may indicate the relative severity of COVID-19 pandemic risk in countries, especially if those countries are financial, economic or even travel hubs in the region.

For tracing the latest COVID-19 pandemic status, we have included the most updated risk scores and risk contributions in http://covid-19-dev.github.io/.

Methods

As presented in Fig. 1, the first step of the proposed pandemic risk assessment system is to construct dynamic pandemic networks using publicly available confirmed COVID-19 cases.

Network statistics. Let $X_{i,t}$ be the number of confirmed COVID-19 cases of country i in day t. Following the literature[9,22], we obtain the daily changes for each country as $Y_{i,t} = \sqrt{X_{i,t}} - \sqrt{X_{i,t-1}}$. To construct dynamic pandemic networks, we calculate $\rho_{ij,t}$, the correlation between country i and country j's daily changes in day t using the observations $(Y_{i,t-k}, Y_{j,t-k})$, for $k = 0, ..., 13$. After that, we construct the pandemic network at time t by defining $A_{ij,t}$, which is the (i, j)th element of an adjacency matrix of the pandemic network at time t:

$$A_{ij,t} = \begin{cases} 1 & \text{if } \rho_{ij,t} > r, \\ 0 & \text{otherwise}. \end{cases} \tag{1}$$

In other words, an edge/link between country i and country j is formed in the pandemic network at time t if $\rho_{ij,t} > r$. Since the earliest report of the number of confirmed cases by the WHO in late January 2020, we have more countries in giving us the data of $X_{i,t}$. To study the statistical properties of the pandemic network from February to December 2020, we start from recording the number of countries having confirmed cases records at time t, denoted by V_t, which is less than or equal to 164 (country names listed in Figs. 5 and 6). Statistically, V_t is the number of vertices of the pandemic network at time t. When a country has zero $Y_{i,s}$ in 14 consecutive days ($s = t - 13, ..., t$), it will be excluded from V_t. Figure 2 shows the pandemic networks at $t = 1$ (4 February 2020), $t = 36$ (11 March 2020) and $t = 67$ (11 April 2020), from which we observe increasing level of connectedness in the pandemic networks from early February to early April[22]. To quantify the network connectedness and to summarize other network properties, we consider four network statistics at time t, E_t (the number of edges), D_t (the edge density), C_t (global clustering coefficient), and AS_t (assortativity). Since we define the edge of the pandemic network by the event that the correlation $\rho_{ij,t}$ exceeds the threshold r, the number of edges, E_t, gives the number of pairs of countries showing a 'common trend' in terms of the confirmed COVID-19 cases in the past 14 days, including time t. In this paper, we follow the literature[9] to set $r = 0.5$. The network density, D_t, defined by

$$D_t = \frac{2E_t}{V_t(V_t - 1)}, \tag{2}$$

is to understand how dense the pandemic network is at time t. The literature[9] presents visual evidence on using E_t to give an early warning signal of the COVID-19 pandemic.

To capture how strong vertices tend to be linked together, we calculate a clustering coefficient of the pandemic network at time t. For vertex i, define its local clustering coefficient at time t as

$$c_{it} = \frac{e_{it}}{\binom{k_{it}}{2}}, \tag{3}$$

where e_{it} is the number of connected pairs among the neighbors of vertex i at time t, and k_{it} is the number of neighbors (or degree) of vertex i at time t. The numerator in Eq. (3), e_{it}, also counts the number of triangles formed by vertex i in the pandemic network at time t, and $\binom{k_{it}}{2}$ is the number of connected triples (a subgraph of three vertices connected by two edges) at time t connecting vertex i. The c_{it} measures the tendency for vertex i to form triangles with its neighbors. The clustering coefficient, denoted by C_t, is a weighted average:

$$C_t = \frac{\sum_{i=1}^{V_t} \binom{k_{it}}{2} c_{it}}{\sum_{i=1}^{V_t} \binom{k_{it}}{2}}. \tag{4}$$

The clustering coefficient C_t measures how strong vertices/countries in the pandemic network at time t are bounded together in clusters. In the extreme case where $C_t = 1$, all pairs vertices/countries are connected by

edges in the pandemic network at time t. In other words, the higher the C_t, the stronger the evidence of global pandemic is revealed in the pandemic network.

$$AS_t = \frac{V_t^{-1}\sum_{i=1}^{V_t}\sum_{j=1}^{V_t}k_{it}k_{jt}I(A_{ij,t}=1) - \left[V_t^{-1}\sum_{i=1}^{V_t}\sum_{j=1}^{V_t}\frac{1}{2}(k_{it}+k_{jt})I(A_{ij,t}=1)\right]^2}{V_t^{-1}\sum_{i=1}^{V_t}\sum_{j=1}^{V_t}\frac{1}{2}(k_{it}^2+k_{jt}^2)I(A_{ij,t}=1) - \left[V_t^{-1}\sum_{i=1}^{V_t}\sum_{j=1}^{V_t}\frac{1}{2}(k_{it}+k_{jt})I(A_{ij,t}=1)\right]^2}, \tag{5}$$

In addition to the network density and the clustering, we also consider the assortativity of networks[31], which is used to describe assortative mixing properties in networks. Specifically, we evaluate the tendency for countries with similar degrees to link together in the pandemic networks. The degree of country i is defined as the number of countries connecting that country at time t and is given by k_{it}. Using the network terminology, k_{it} is the number of edges incident to vertex/country i. In epidemiological perspectives, the risk of individuals' infection may be related to the assortativity (high degree vertices tend to link with other vertices with high degrees) in social networks[24]. In this paper, we apply this assortative mixing concept to the pandemic networks constructed to understand how the assortative mixing properties change over time during the COVID-19 pandemic. Statistically speaking, we calculate an assortativity coefficient which measures the correlation of the degress of the vertices in the pandemic networks. At time t, the assortativity coefficient is defined by Eq. (5) where $I(E)$ is an indicator function whose value is equal to one if the event E is true and zero otherwise. We attempt to learn any insight from this assortativity coefficient on possible early warning signals of a global pandemic like COVID-19.

Pandemic risk scores. In this paper, we propose two pandemic risk scores defined by the dynamic pandemic networks represented by $A_{ij,t}$ in Eq. (1), the adjacency matrix at time t. Specifically, we aggregate the link information in $A_{ij,t}$, susceptible population sizes (or population at risk) of each country, and the number of confirmed cases, $X_{i,t}$, to construct two scores which help us to assess the potential transmission risk across countries, and the severity of the global pandemic risk due to COVID-19. Define the first risk score as the preparedness risk score (PRS)

$$S_{1t} = \omega_t^\top A_t \omega_t, \tag{6}$$

where ω_t is the vector of population size of each country (based on the World Bank population figures in 2018) subtracted by the total number of confirmed cases in each country up to time t. In other words, the ith element of ω_t, denoted by ω_{it}, represents the population of country i at risk or susceptible at time t. Since $S_{1t} = \sum_{i=1}^{V_t}\sum_{j=1}^{V_t}\omega_{it}A_{ij,t}\omega_{jt}$, S_{1t} in Eq. (6) counts the total number of possible interactions of susceptible population contributed from all pairs of countries which are linked together at time t. This PRS accounts for the risk of asymptomatic transmission or presymptomatic transmission[32] due to possible interaction between people in the two countries. These two kinds of transmission are usually hidden in the population and is useful to quantify the pandemic risk especially when the transmission between two countries cannot be completely stopped by travel restrictions or other infection prevention policies. Therefore, the risk score, S_{1t}, can help us to project the severity of the COVID-19 outbreak for preparedness if existing travel restrictions or lockdown schemes are released.

We define the second risk score as the severity risk score (SRS)

$$S_{2t} = \omega_t^\top A_t n_t, \tag{7}$$

where n_t is the vector of the total number of currently infected people at time t. In other words, the ith element of n_t, denoted by n_{it} is calculated as $n_{it} = \sum_{s=1}^{t} X_{i,s} - AR_{i,t} - AD_{i,t}$, where $AR_{i,t}$ and $AD_{i,t}$ are the accumulated number of recovered cases and accumulated number of deaths due to COVID-19 in country i at time t. It is easy to see that $S_{2t} = \sum_{i=1}^{V_t}\sum_{j=1}^{V_t}\omega_{it}A_{ij,t}n_{jt}$. The rationale behind S_{2t} is to count the number of possible interactions between currently infected cases of one country and all people at risk in another country, if the two countries are linked together. The product of $\omega_{it}n_{jt}$ is similar to the classical SIR model[7], where the number of susceptible people at time t multiplied with the number of infected people at time t is used to determine the rate of change of the number of susceptible people. This SRS accounts for presymptomatic transmission due to possible interaction between susceptible population in one country and currently infected cases in another country before the confirmed cases are identified and forced-isolated or quarantined. The SRS, S_{2t}, can help assess the severity of the COVID-19 due to the current level of infections and plan for outbreak control measures. We present this step in computing the PRS and SRS in the work flow diagram in Fig. 1.

As in the literature[33], we can define the risk contribution by country i on the PRS as

$$f_{it} = \frac{\omega_{it}\frac{\partial\sqrt{S_{1t}}}{\partial\omega_{it}}}{\sqrt{S_{1t}}} = \frac{\omega_t^\top A_t \mathbf{1}_i \mathbf{1}_i^\top \omega_t}{\omega_t^\top A_t \omega_t}, \tag{8}$$

where $\mathbf{1}_i$ is a column vector with the i-th entry equal to 1 and 0 otherwise. Using the Euler rule, we can show that $\sum_i f_{it} = 1$. From Eq. (8), f_{it} is approximately equal to the ratio of the percentage change in $\sqrt{S_{1t}}$ over the percentage change in ω_{it}. If a particular country i has high f_{it}, reducing the mobility of the people (at risk) in that country may lead to substantial reduction in S_{1t}, thereby lowering the global pandemic risk. Similarly, we can define the risk contribution by country i on the SRS as

$$\frac{n_{it}\frac{\partial S_{2t}}{\partial n_{it}}}{S_{2t}} \approx \frac{n_{it}\sum_{j=1}^{V_t}\omega_j A_{jl}}{S_{2t}} = g_{it}. \tag{9}$$
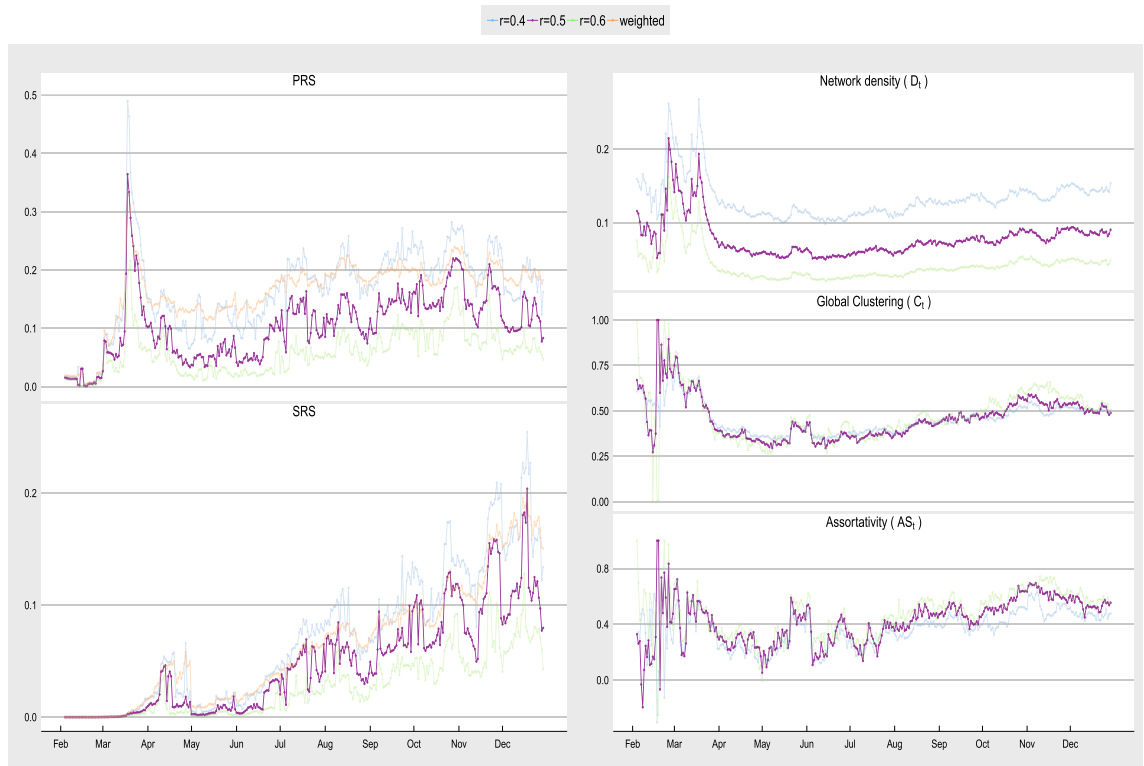
Figure 7. The left panel shows the PRS and SRS, and the right panel shows the network density, clustering coefficient and assortativity for $r = 0.4, 0.5$ and 0.6.

Again, we can show that $\sum_i g_{it} = 1$. Equation (9) implies that g_{it} is approximately equal to the ratio of the percentage change in S_{2t} over the percentage change in n_{it}. If a particular country i has high g_{it}, quarantining currently infected cases and enhancing social distancing measures as best as we can help 'block' the interaction of infected people with people at risk, and may lead to substantial reduction in S_{2t}, thereby achieving better control on the global pandemic. In summary, for outbreak prediction or prevention, S_{1t} is recommended for policymakers to make reference to. For outbreak control, S_{2t} is more meaningful.

The risk contributions f_{it} and g_{it} with respect to the PRS and SRS can be computed on the global, regional, and country levels (see Fig. 1). For active pandemic risk monitoring, the most updated PRS, SRS, f_{it}, and g_{it} can be accessed in http://covid-19-dev.github.io/.

Sensitivity analysis

The construction of the dynamic pandemic network in Eq. (1) requires the specification of the threshold value r. To perform sensitivity analysis on the choice of r, we study the pandemic network properties for $r = 0.4, 0.5$ and 0.6. Figure 7 shows the PRS, SRS, network density, clustering coefficient and assortativity based on the pandemic networks constructed using $r = 0.4, 0.5$ and 0.6. When r is smaller, we expect to see more connections and so the three network density time series in Fig. 7 differ with the highest one appears at $r = 0.4$. Although the three network time series have different levels, they exhibit very similar trends. In addition, both the clustering coefficient and assortativity are insensitive to the choice of r, implying that major network properties regarding pandemic network connectedness are not sensitive to the choice of the threshold r if it falls in 0.4 to 0.6. A reviewer suggests that the adjacency matrix can be weighted to reflect the different level of connection between countries. Based on this suggestion, we also compare the PRS and SRS with the adjacency matrix defined by $A_{ij,t} = \rho_{ij,t}$ if $\rho_{ij,t} > 0$ and $A_{ij,t} = 0$ otherwise. We can see from Fig. 7 that all four PRS and SRS show similar temporal patterns, indicating that the A_t defined by different r and the A_t defined by the weight $\rho_{ij,t}$ give consistent pandemic risk assessment based on the PRS and SRS. The above sensitivity analysis demonstrates the robustness of our proposed pandemic network methodology, which enables users to pick their own thresholds without changing major conclusions for pandemic risk prediction and assessment of COVID-19.

There is also a comment that the SRS is better calculated with the use of the number of local COVID-19 cases rather than the total confirmed cases. This is a good point. However, most countries do not provide local case information. Since the percentage of imported cases is relatively small after extensive travel restrictions after March 2020[34], the imported cases should contribute very little in the susceptible population and so the general trend of the SRS will likely be the same even though we replace the number of confirmed cases by the number of the local cases. Therefore, we keep the definition of the SRS in Eq. (7) in our application to the COVID-19 pandemic. Another related issue in the calculation of the SRS is the availability of the number of recovered cases. As WHO only contains official information of the number of COVID-19 confirmed cases and the number of death
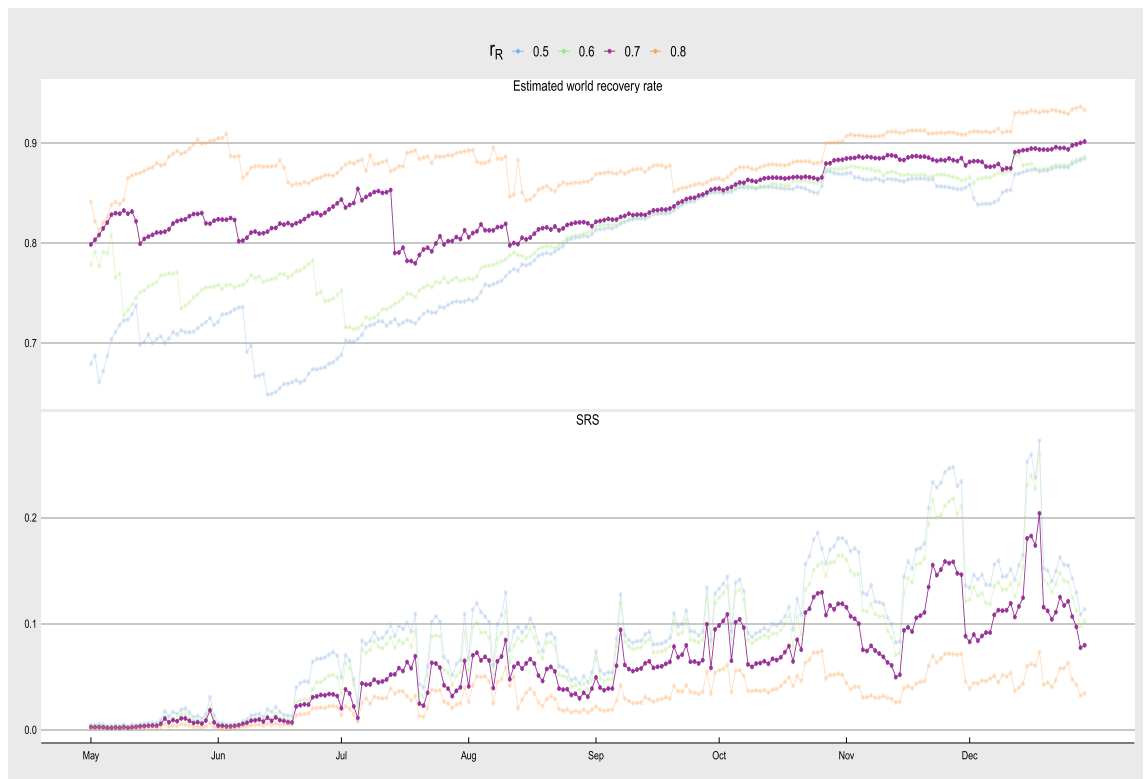
Figure 8. The top panel shows the estimated world recovery rate based on $r_R = 0.5$ to 0.8. The bottom panel shows the SRS based on the imputed number of recovered cases with $r_R = 0.5$ to 0.8.

due to COVID-19, the number of recovered cases can be missing or under-reported in some countries[35–37]. To deal with incomplete recovered case information, we perform additional analysis to extract possible tendency of the world COVID-19 recovery rate over time. We first set a threshold r_R which classifies whether the number of recovered cases is under-reported or missing. Define $R_{i,t}^{(R)}$ as the reported recovery rate of country i at time t, i.e. the total reported number of recovered cases up to time t divided by the total number of confirmed cases up to time t. We estimate the world recovery rate by filtering out those countries which have $R_{i,t}^{(R)} < r_R$, and calculate

$$R_{w,t} = \frac{\sum_{i:R_{i,t}^{(R)} \geq r_R} AR_{i,t}^{(R)}}{\sum_{i:R_{i,t}^{(R)} \geq r_R} \sum_{s=1}^{t} X_{i,s}},$$

(10)

where $AR_{i,t}^{(R)}$ is the accumulated number of recovered cases of country i reported up time t. Therefore, $R_{w,t}$ is an estimated world recovery rate at time t based on those countries whose recovery rate at time t exceeds r_R. Figure 8 presents the estimated world recovery rate based on Eq. (10) with $r_R = 0.5$ to 0.8. The four plots in the figure show an increasing trend in the recovery rate from May 2020 to December 2020. In mid December 2020, the four estimated recovery rates based on different r_R are consistent that they seem to converge to around 0.9, which is in line with some published recovery rates[38–40].

For countries whose $R_{i,t}^{(R)} < r_R$, we regard their number of recovered cases either under-reported or missing. Then, we do imputation for the countries with $R_{i,t}^{(R)} < r_R$ by taking $AR_{i,t} = R_{w,t} \times \sum_{s=1}^{t} X_{i,s}$. That is, we substitute the under-reported or missing recovery rates at time t by $R_{w,t}$. For countries whose $R_{i,t}^{(R)} \geq r_R$, we take $AR_{i,t} = AR_{i,t}^{(R)}$, that is, we regard the accumulated number of recovered cases reported as official valid values to be adopted in the SRS calculation. The bottom panel of Fig. 8 gives the SRS based on the four r_R. Again, the four SRS time series pick up a similar trend that we are able to identify local peaks in early October, late October, late November and mid December 2020. We adopt $r_R = 0.7$ in Figs. 4 and 6 as the time series of $R_{w,t}$ is relatively more stable when $r_R = 0.7$. Given that the SRS is to capture the relative severity of the pandemic risk for now-casting the COVID-19 pandemic, the robustness in SRS to the choice of r_R gives us flexibility in adjusting r_R in real applications.

Data availibility
The authors confirm that the real data sets used in this study are freely available to public at the time of writing by accessing https://github.com/CSSEGISandData/COVID-19 and https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed on 31 December 2020).

Code availability
All data and codes are available in the Zenodo webpage: http://doi.org/10.5281/zenodo.4408746[41].

References

1. WHO. Critical preparedness, readiness and response actions for COVID-19: interim guidance, 24 June 2020. Tech. Rep., World Health Organization (2020).
2. Gilbert, M. *et al.* Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet* **395**, 871–877 (2020).
3. Gatto, M. *et al.* Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl. Acad. Sci.* **117**, 10484–10491 (2020).
4. Moghadas, S. M. *et al.* Projecting hospital utilization during the COVID-19 outbreaks in the united states. *Proc. Natl. Acad. Sci.* **117**, 9122–9126 (2020).
5. Wells, C. R. *et al.* Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci.* **117**, 7504–7509 (2020).
6. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **395**, 689–697 (2020).
7. Weiss, H. H. The SIR model and the foundations of public health. *MATerials MATematics* **0001–17**, (2013).
8. Depoux, A. *et al.* The pandemic of social media panic travels faster than the COVID-19 outbreak. *J. Travel Med.* https://doi.org/10.1093/jtm/taaa031 *(2020)*.
9. Chu, A. M. Y., Tiwari, A. & So, M. K. P. Detecting early signals of COVID-19 global pandemic from network density. *J. Travel Med.* https://doi.org/10.1093/jtm/taaa084 *(2020)*.
10. Jewell, N. P., Lewnard, J. A. & Jewell, B. L. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA* **323**, 1893–1894 (2020).
11. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet* **395**, 931–934 (2020).
12. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400 (2020).
13. Newman, M. *Networks* (Oxford University Press, Oxford, 2018).
14. Newman, M. E. Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci.* **101**, 5200–5205 (2004).
15. Diebold, F. X. & Yılmaz, K. On the network topology of variance decompositions: measuring the connectedness of financial firms. *J. Econ.* **182**, 119–134 (2014).
16. Elliott, M., Golub, B. & Jackson, M. O. Financial networks and contagion. *Am. Econ. Rev.* **104**, 3115–53 (2014).
17. Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**, 2015–2020 (2006).
18. Bansal, S., Read, J., Pourbohloul, B. & Meyers, L. A. The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dyn.* **4**, 478–489 (2010).
19. Danon, L. *et al.* Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* **2011**, 284909 (2011).
20. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLOS Comput. Biol.* **4**, e1000117 (2008).
21. Vega, D. & Magnani, M. Foundations of temporal text networks. *Appl. Netw. Sci.* **3**, 25 (2018).
22. So, M. K. P., Tiwari, A., Chu, A. M. Y., Tsang, J. T. Y. & Chan, J. N. L. Visualizing COVID-19 pandemic risk through network connectedness. *Int. J. Infect. Dis.* **96**, 558–561 (2020).
23. Daudin, J.-J., Picard, F. & Robin, S. A mixture model for random graphs. *Stat. Comput.* **18**, 173–183 (2008).
24. Christley, R. M. *et al.* Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.* **162**, 1024–1031 (2005).
25. Sardar, T., Nadim, S. S., Rana, S. & Chattopadhyay, J. Assessment of lockdown effect in some states and overall India: a predictive mathematical study on COVID-19 outbreak. *Chaos Solitons Fractals* **139**, 110078 (2020).
26. Furukawa, N. W., Brooks, J. T. & Sobel, J. Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. *Emerging Infectious Diseases* **26**(7), e201595 (2020).
27. Lauer, M. S. *et al.* Impaired chronotropic response to exercise stress testing as a predictor of mortality. *JAMA* **281**, 524–529 (1999).
28. Johnson, H. C. *et al.* Potential scenarios for the progression of a Covid-19 epidemic in the European union and the European economic area, March 2020. *Eurosurveillance* **25**, 2000202 (2020).
29. Remuzzi, A. & Remuzzi, G. COVID-19 and Italy: what next? *The Lancet* **395**(10231), 1225–1228 (2020).
30. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
31. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
32. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**(6491), eabb6936 (2020).
33. So, M. K. P., Chan, T. W. C. & Chu, A. M. Y. Efficient estimation of high-dimensional dynamic covariance by risk factor mapping: applications for financial risk management. *J. Econom.* https://doi.org/10.1016/j.jeconom.2020.04.040 (2020).
34. Jefferies, S. *et al.* COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study. *Lancet Public Health* **5**, e612–e623 (2020).
35. Dyer, O. COVID-19: US states are not reporting vital data, says former CDC chief. *BMJ* **370**, m2993 (2020).
36. Project, C. T. Data faq: Why have you stopped reporting national recoveries? (2020). https://covidtracking.com/about-data/faq#why-have-you-stopped-reporting-national-recoveries.
37. DEVEX. What you need to know about 'active' and 'recovered' COVID-19 cases. (2020). https://www.devex.com/news/what-you-need-to-know-about-active-and-recovered-covid-19-cases-98525.
38. DDNews. India's COVID-19 recovery rate improves to 93.58 percent. (2020). http://ddnews.gov.in/national/indias-covid-19-recovery-rate-improves-9358.
39. SANews. Sa records close to 92 percent COVID-19 recovery rate. (2020). https://www.sanews.gov.za/south-africa/sa-records-close-92-covid-19-recovery-rate.
40. Moaje, M. Doh reports 90.1 percent COVID-19 recovery rate. (2020). https://www.pna.gov.ph/articles/1122560.
41. So, M. K. P., Chu, A. M. Y., Tiwari, A. & Chan, J. N. L. Source code for: On topological properties of covid-19: predicting and assessing pandemic risk with network statistics (version v1.0). zenodo. (2021). https://doi.org/10.5281/zenodo.4408746.

Acknowledgements

Author contributions

So, M.K.P. and Chu, A.M.Y. conceptualized the study. M.K.P.S. and J.N.L.C., collected and analyzed the data. M.K.P.S., A.M.Y.C. and A.T. interpreted the results. M.K.P.S. and A.M.Y.C. drafted the manuscript. M.K.P.S., A.M.Y.C. and A.T. finalized the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.K.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.