

# Identifying kidney trade networks using web scraping data

Meng-Hao Li,<sup>1</sup> Abu Bakkar Siddique ,<sup>1</sup> Brian Wilson,<sup>1</sup> Amit Patel,<sup>2</sup> Hadi El-Amine,<sup>3</sup> Naoru Koizumi <sup>1</sup>

**To cite:** Li M-H, Siddique AB, Wilson B, *et al.* Identifying kidney trade networks using web scraping data. *BMJ Global Health* 2022;**7**:e009803. doi:10.1136/bmjgh-2022-009803

**Handling editor** Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2022-009803>).

Received 13 June 2022  
Accepted 29 August 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Schar School of Policy and Government, George Mason University, Arlington, Virginia, USA

<sup>2</sup>Department of Public Policy and Public Affairs, University of Massachusetts Boston, Boston, Massachusetts, USA

<sup>3</sup>Systems Engineering and Operations Research Department, George Mason University, Fairfax, Virginia, USA

**Correspondence to**  
Dr Naoru Koizumi;  
[nkoizumi@gmu.edu](mailto:nkoizumi@gmu.edu)

## ABSTRACT

Kidney trade has been on the rise despite the domestic and international law enforcement aiming to protect the vulnerable population from potential exploitation. Regional hubs are emerging in several parts of the world including South Asia, Central America, the Middle East and East Asia. Kidney trade networks reported in these hot spots are often complex systems involving several players such as buyers, sellers and surgery countries operating across international borders so that they can bypass domestic laws in sellers and buyers' countries. The exact patterns of the country networks are, however, largely unknown due to the lack of a systematic approach to collect the data. Most of the kidney trade information is currently available in the form of case studies, court materials and news articles or reports, and no comprehensive database exists at this time. The present study thus explored online newspaper scraping to systematically collect 10 419 news articles from 24 major English newspapers in South Asia (January 2016 to May 2019) and build transnational kidney trade networks at the country level. Additionally, this study applied text mining techniques to extract words from each news article and developed machine learning algorithms to identify kidney trade and non-kidney trade news articles. Our findings suggest that online newspaper scraping coupled with the machine learning method is a promising approach to compile such data, especially in the dire shortage of empirical data.

## INTRODUCTION

International health authorities and law enforcement agencies have taken several initiatives against organ trade in the last decade. In 2004, the World Health Assembly urged actions against the practice of human organ trade.<sup>1</sup> Following this, the Declaration of Istanbul Custodian Group has become the principal international agency in 2008 to combat organ trade and transplant tourism in coordination with The Transplantation Society.<sup>2</sup> More recently, in 2017, with the directive of Pope Francis, the Vatican arranged a summit meeting on organ sales and transplant tourism to combat this practice worldwide.<sup>3</sup> These movements are partially attributable to the recent recognition that organ trade frequently occurs as part of transnational

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The operation of transnational kidney trade consists of multiple agents, typically involving kidney sellers, buyers and transplant service providers, but no systematic approach has been explored to identify and investigate the pattern of such networks.

## WHAT THIS STUDY ADDS

⇒ Collecting data from news articles via web scraping can be a promising approach to study kidney trade market and to study formation of transnational kidney trade networks.

⇒ This study applied machine learning algorithms to automatically classify kidney trade news articles.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The analytical tools and process offered by this study could be applied to examine the structure and attributes of transnational organ trade networks.

⇒ A better understanding of transnational organ trade networks would assist law enforcement agencies around the world in making more informed decisions.

organised crime, as evidenced in Middle Eastern criminal organisations, which exploit refugees who sell their kidneys to pay for their passage to Europe.<sup>4-6</sup>

The operation of organ trade is complex. It consists of multiple transnational agents, typically involving organ sellers (often poor, young and healthy), buyers (affluent but desperate), transplant service providers (surgeons, hospitals, labs), brokers (who often provide similar services to other illicit trades) and financial institutions that transfer money from buyers to other agents. For instance, the kidney trade network revealed in Costa Rica in 2013 comprised Costa Rican donors and medical service providers, Israeli and European buyers/patients, a Greek broker who maintained the façade of a pizzeria, a Costa Rican nephrologist who became a broker and a Ukrainian crime organisation.<sup>7</sup> Also in 2011, one of the largest kidney trade

networks in history was found in Bangladesh. A group of brokers led by one kidney-seller-turned-broker arranged at least 43 villagers to sell their kidneys. Many buyers were Bangladesh-descent immigrants living in the Middle East, Europe and North America, while the institutions performing surgeries were in India, Pakistan, Singapore and Thailand.<sup>8</sup>

Though the intricate interconnectivity among agents across multiple countries has been reported, no quantitative approach has been explored to identify and investigate the pattern of these transnational organ trade networks. Most of these networks are revealed and discussed individually as case studies and reports, news articles and other media. While the effort to compile the information of organ trade networks exists including the European Union-supported 'HoTT project', such endeavour takes a form of a repository of limited number of case studies and reports, rather than a database recording the characteristics of the organ trade networks.<sup>9</sup> As such, the present study sought to develop an exploratory approach to gather data on kidney trade networks reported in well-circulated, reliable online newspapers. In this endeavour, we focused on online newspapers circulated in South Asia, the region known as a major organ trade hub for more than two decades.<sup>10–13</sup> We focused on kidneys as opposed to other organs since kidney sales encompass most cases of organ trade involving living sellers. In total, we accessed 24 online newspapers published between January 2016 and May 2019 to build transnational kidney trade networks, consisting of: (1) seller's nationality; (2) buyer's nationality; and (3) location of the surgery.

To compile the data, we first applied a standard web scraping method to collect news articles that are potentially reporting kidney trade. A team of researchers then manually assessed whether each article is in fact reporting transnational kidney trade. We then applied conventional machine learning (ML) algorithms to automate the classification of the articles into two sets: (1) kidney trade article set and (2) non-kidney trade article set. The final ML algorithm was fairly effective with the sensitivity and specificity levels of 74% and 80%, respectively. The result indicates that ML-based web scraping could be a useful tool to collect the data. During the manual evaluation of news articles, we recorded buyers and sellers' nationalities as well as the countries of surgeries for visualisation of South Asian transnational kidney trade networks.

This paper is structured in the following fashion. The next section presents the data sources (section 2), section 3 describes the methods applied for the news article collection. Section 4 presents the results of the web scraping, ML algorithm and kidney trade network visualisation. Sections 5 and 6 provide discussion and conclusion, respectively.

## DATA SOURCE

Given the large number of English newspapers circulated in South Asia, we focused on the two regions known

as the hot spots of kidney trade within South Asia: (1) West Bengal in India, Bangladesh and Nepal (hereafter denoted as 'Region I'); and (2) Pakistan, which borders India and Afghanistan (hereafter denoted as 'Region II'). We focused on West Bengal in India as the region is likely to be involved in transnational kidney trade, particularly with Bangladesh and Nepal, the two countries known as hot spots.<sup>8 14 15</sup> Pakistan was selected because it is another known hot spot in the region,<sup>16</sup> and also because the kidney trade cases reported in Pakistan tend not to be covered by Bangladeshi newspapers. While Pakistan borders with state or Rajasthan, Punjab, Gujarat, and Jammu and Kashmir in India, we did not include newspapers from those Indian regions as we suspected a low volume of illicit trade between India and Pakistan due to their historically tensed relationships and heightened border security on both sides, as well as thorough background checks of visitors' visa applications, which make it hard to operate such illicit networks.

We identified 12 well-circulated newspapers from each region, totalling 24 English newspapers that are considered trustworthy by the locals (online supplemental table A1). These selected newspapers collectively covered more than 90% of the total English readership in each region.<sup>17 18</sup> All selected papers were available online for free of charge, making web scraping possible without additional administrative and financial burdens. We did not include newspapers in local languages for this exploratory study, following the recommendation of local experts who commented that English and local languages have significant overlaps. Most articles were on the front page section, featuring trafficking rings uncovered by local law enforcement agencies. A typical length of the article was around 700 words with a graphic (eg, ref 19). Some representative texts extracted from these articles include: '... the heinous crime of kidney smuggling to India from Nepal was revealed' (2018); and 'All participants were ... at the district headquarters in Dhulikhel (Nepal)... "B"roker had forced him to sell his kidney after taking him to New Delhi...' (2016).

## METHODS

### Patient and public involvement

No patients or the public were involved in the design of our research.

### Data preprocessing

A standard approach was taken to web scrape the articles. First, key search terms were identified via trial and error, which went in tandem with a series of discussions with several kidney trafficking experts. The final key terms included: kidney+one of the following terms: 'trafficking'; 'trade'; 'buying'; 'selling'; 'market'; 'bazaar'; 'sales'; and 'illegal transplant'. Fuzzy matching was applied using the format—'site: ~ search term' on google.com. We used Python libraries including Selenium, NLTK, GeoText, Requests and BeautifulSoup to execute

the search. From the initial set of web scraped articles, we extracted article fragments that contained location identifiers (cities, states, countries) using GeoText, as well as any variations of the terms, buying, selling and surgery. Next, four reviewers (two faculty members and two PhD students) went through each of the articles to manually assess whether the article indeed reported: (1) kidney trade case; (2) the nationalities of kidney sellers and/or buyers; and (3) the countries where surgeries took place. The data on the nationalities and surgery locations were visualised as transnational kidney trade networks using Gephi, an open-source software for visualising networks. A total of 10 419 news articles were extracted via web scraping, among which 276 reported kidney trade.

An ML algorithm was trained to classify each article as either a ‘kidney trade article’ or a ‘non-kidney trade article’. One of the major tasks in the algorithm development was to identify a set of keywords that most effectively classify articles into these two groups. To identify such keywords, we first encoded all the data (ie, raw texts of the 10 419 news articles) into the UTF-8 standard and converted all the words into lower case. We then removed special symbols (ie, ^, <, \*), numbers and stop words (eg, also, in, if) in the data using System for the Mechanical Analysis and Retrieval of Text.<sup>20</sup> We also stripped extra whitespace to collapse multiple whitespace characters to a single blank and stemmed the words using Porter’s stemming algorithm.<sup>21</sup> From the resulting set of the words, we identified those words that uniquely appeared in kidney trade articles and non-kidney trade articles. These words were then used as the input variables to train the ML algorithm.

Given that the data set mostly consisted of non-kidney trade articles (the proportion of kidney trade articles was 276 out of 10 419, a mere 2.6% of the total), the

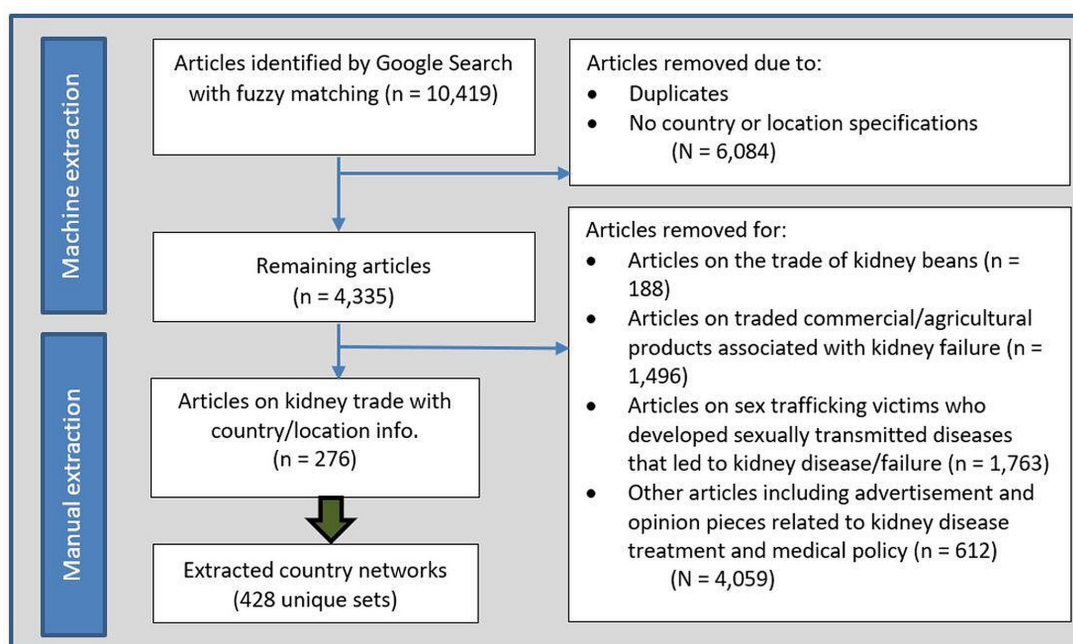
imbalanced response variable was adjusted using the synthetic minority oversampling technique (SMOTE), a technique used to account for the imbalance in the data set. Finally, we applied Random Forest and Neural Network algorithms, two well-established ML algorithms, to build classification models, and then split data into 10-folds to validate both models. The performance of the two algorithms was evaluated using area under the receiver operating characteristic curve (AUC). We used two R packages, tm and caret, to implement data preprocessing and ML. The tm package was used to extract words from news articles and manage the word cleaning process. The caret package was used to implement SMOTE, Random Forest and Neural Network algorithms, model cross-validation and model evaluation.<sup>22 23</sup>

## RESULTS

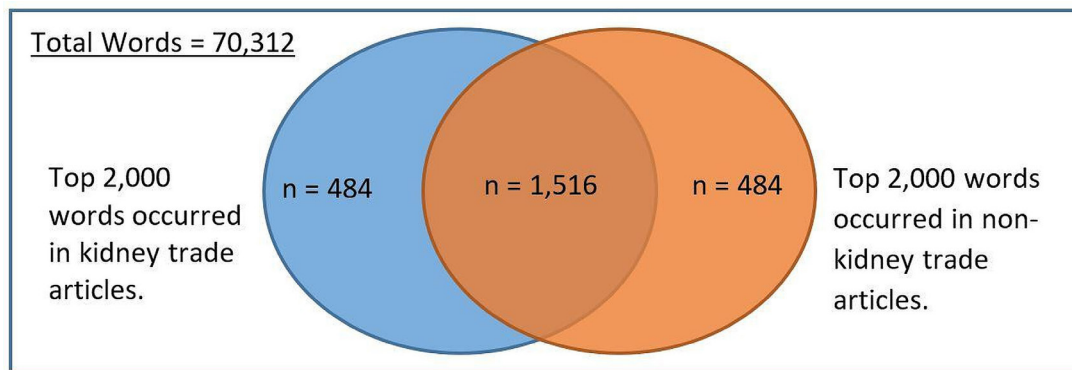
This section comprised: (1) the results from the web scraping; (2) the results from the ML analysis; and (3) the visualised country-level kidney trade networks.

### Web scraping results

A total of 10 419 newspaper articles were collected via web scraping. We first eliminated duplicates and the articles without any location identifiers by recoding in Python and using GeoText, respectively. This process eliminated 6084 articles, leaving 4335 articles. The four reviewers then manually evaluated 4335 articles to further filter irrelevant articles and to extract kidney trade network information. This process eliminated 4059 additional articles, leaving 276 unique articles reporting kidney trade. [Figure 1](#) presents the flow chart of the article and network extraction steps. The 4059 eliminated articles



**Figure 1** Flow chart of article collection to network extraction.



**Figure 2** Word selection for machine learning (ML) algorithm development.

had the following major reasons for erroneously being identified as the articles reporting kidney trade:

- ▶ The articles on the *trade of kidney* beans (n=188).
- ▶ The articles on *traded* commercial/agricultural products (eg, pesticide, drug, hair dye) that are associated with *kidney* failure (n=1496).
- ▶ The articles on *sex trafficking* of which victims tend to get contracted with sexually transmitted diseases that led to *kidney* disease/failure (n=1763).
- ▶ Other miscellaneous articles including those that had an advertisement of *kidney* disease treatment and other related *sales*, such as dialysis services and transplant, on the sidebars (n=612).

### ML application for article classification

The ML algorithm was trained to automate the identification of the 276 kidney trade articles from the pool of the 10419 web scraped articles. For the algorithm development, we first performed data cleaning described in the Methods section to extract the words that are useful in classifying the articles. The data cleaning identified 70312 distinct words in the 10419 articles. Among these words, we listed 2000 words with the highest word counts in each of the kidney trade (n=276) and non-kidney trade (n=10143) article sets. The word lists were then compared to identify 484 words that are uniquely associated with the kidney trade and non-kidney trade article sets, respectively (figure 2).

Among the 484 words, we manually identified 104 words that are seemingly more directly and explicitly associated with kidney trade. These 104 words were then used as the input variables in the ML models. This filtering process narrowed down the keywords used to effectively predict the kidney trade articles. The words are listed below:

*Included words*—abroad, afford, broker, cash, caught, clinic, imprison, miser, physician, remand, sponsor.

*Excluded words*—adjust, adopt, brother, complain, difficult, figure, hotel, mention, neighbor, plain, relationship, scare, undergo, valley, want.

Random Forest and Neural Network with Principal Component Analysis algorithms were applied to develop training models. The final results indicated that the Neural Network algorithm (AUC=0.774) outperforms the Random Forest algorithm (AUC=0.723). The Neural

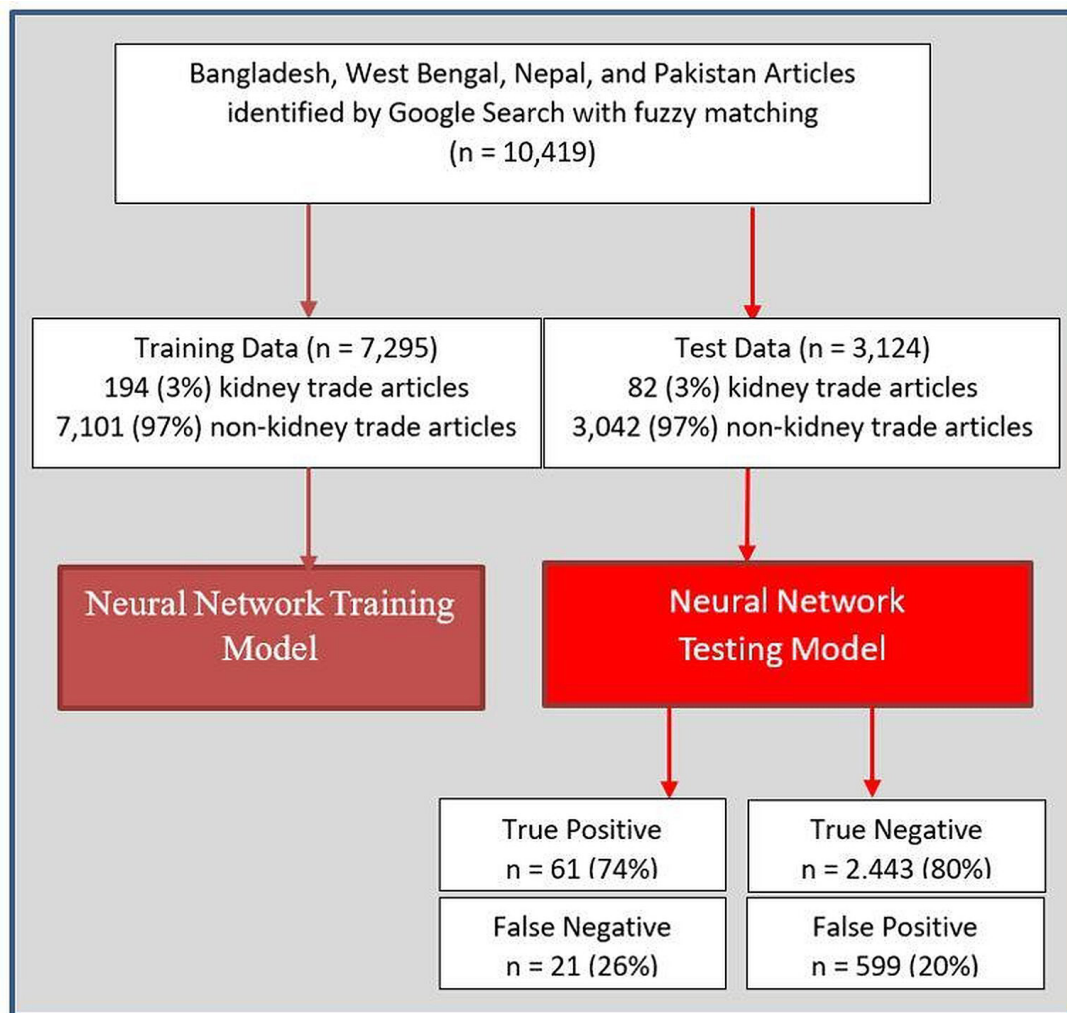
Network algorithm was successful in identifying 61 out of 82 (ie, 74% of) kidney trade articles (ie, sensitivity) and 2443 out of 3042 (ie, 80% of) non-kidney trade articles (ie, specificity). The process of the Neural Network algorithm development and the results are shown in figure 3.

In the process of running the Neural Network model, we ranked these 104 words by the level of contribution in differentiating the kidney trade and non-kidney trade articles.<sup>24</sup> Of those, ‘racket’, ‘tissue’, ‘reluctant’, ‘license’ and ‘poverty’ were the five most important keywords. Accordingly, using these keywords in addition to the aforementioned search terms (kidney+one of the following terms: ‘trafficking’; ‘trade’; ‘buying’; ‘selling’; ‘market’; ‘bazaar’; ‘sales’; and ‘illegal transplant’) could make the web scraping of kidney trade articles more efficient.

### Transnational kidney trade networks

In the final 276 articles reporting kidney trade, we identified 428 unique mentions of transnational kidney trade networks consisting of buyer, seller and surgery county information. Figures 4 and 5 present the extracted transnational kidney trade networks in Region I (Bangladesh, West Bengal and Nepal) and in Region II (Pakistan), respectively. In both figures, buyer countries are shown in blue while seller and surgery countries are shown in red and green, respectively. The size of the circles represents the number of times that the country reported as a seller (red), buyer (blue) or/and surgery (green) country while the width of the lines represents the number of connections reported in the articles (eg, a buyer from country A travelled to country B to receive a surgery).

The thick lines between India (green), Nepal (red) and India (blue) seen in the Region I networks (figure 4) indicate that a large proportion of surgeries that took place in India involved Nepalese sellers and Indian buyers. Similarly, the thick lines connecting Bangladesh (blue), Bangladesh (red) and India (green) indicate that a large proportion of surgeries for the sales between Bangladeshi sellers and Bangladeshi buyers was performed in India. It also appears that a smaller proportion of Bangladeshi seller–buyer pairs received surgeries in Bangladesh (lines connecting blue, red and green circles for Bangladesh), and even a smaller portion in Singapore and Vietnam.



**Figure 3** Text classification process.

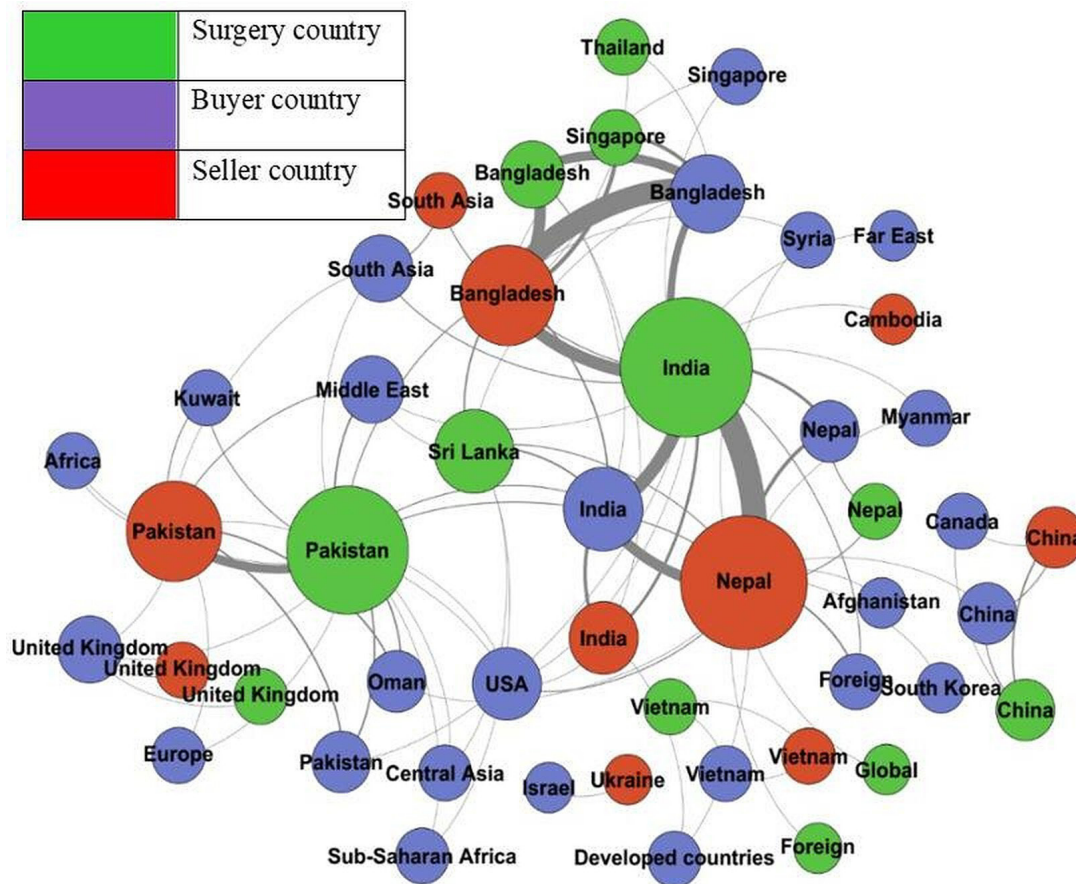
Overall, the Region I networks revealed that a majority of the reported surgeries took place in India, followed by Pakistan, Singapore, Sri Lanka, China, Bangladesh, Vietnam and Thailand, as shown by the relatively large-size green circles in the figure. The sellers were mainly from Nepal, Bangladesh, Pakistan and India, followed by China, Cambodia and Vietnam. A large proportion of Nepalese kidney buyers were Indians, while buyers of Pakistani kidneys were found to be from the Middle East. In contrast, Bangladeshi kidney buyers tended to rely more on domestic sellers. In general, buyers were found to be from all over the world, although many were from South Asia. The other buyers were from the Middle Eastern, Gulf, European and African countries. In some cases, the nationalities of the buyers were not disclosed, and these cases were recorded and shown as ‘foreign’, ‘Far East’ and ‘developed countries’.

The Region II networks presented in [figure 5](#) show mostly confirmatory patterns, demonstrating that Pakistan has mainly served as both surgery and seller country. As seen in [figure 4](#), their reported buyers were overwhelmingly from Middle Eastern and Gulf countries. The major surgery sites were similar to those found in

Region I networks, including India, Pakistan, Sri Lanka, Singapore and China. Major seller countries included, as in Region I networks, India, Pakistan, Bangladesh, Nepal, Cambodia and China. As was the case for Region I, some of the buyers’ nationalities were not disclosed in the articles, and these cases are denoted as ‘foreign’ or using respective continents in [figure 5](#).

## DISCUSSION

The transnational kidney trade networks revealed in the present study had both expected and unexpected results. Countries that were uniquely associated with the nationalities of sellers were the countries with a relatively lower gross domestic product per capita in the region. Several news articles that we manually assessed also reported that poverty and debt repayments are the key reasons that make many people in these countries sell their kidneys. In particular, the networks found in Region I indicated that Nepal and Cambodia mainly supplied sellers and were unlikely to perform surgeries or involve local buyers. This most likely reflects that both countries have limited transplant capacity compared with other countries in



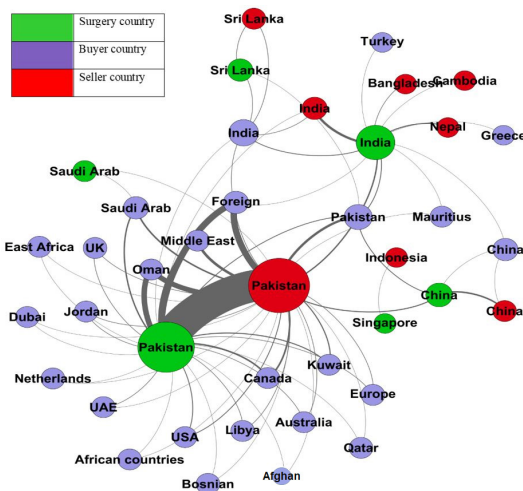
**Figure 4** Kidney trade networks using data from Region I news articles.

the region such as India and Pakistan. At the same time, economic poverty and debt burdens of microcredits likely caused many people in these countries to sell their kidneys as documented in earlier literature.<sup>8 10 11 25</sup>

The key motivation of transnational kidney trafficking is to avoid the legal constraints in the respective countries. Countries with loose legal and procedural frameworks and a reasonable medical infrastructure often drive the transplant tourists' demand, such as Pakistan.<sup>26</sup> While

most countries included in the present study have a legal ban on any kind of monetary transaction between kidney donors and recipients for their own citizens, such a legal ban has not been applied or enforced for foreign citizens. For example, Indian and Nepalese laws ban the sale of kidneys and require that all living donors are conducted for altruistic and not monetary reasons.<sup>27 28</sup> However, a large proportion of surgeries seem to be performed in India involving Nepalese sellers and Indian buyers. This is because the law enforcement institutions do not attempt to verify donor–recipient relationships if donors or buyers are from other countries. Knowing this loophole, professional brokers arrange fake passports and forged documents showing that the donors are relatives of the recipients, and advise sellers to hide their identities so that immigration officers do not reject the case.<sup>8</sup> Indeed, the system of the countries hosting illegal transplants has neither the capacity nor the incentive to verify whether these documents are authentic or not. Here, establishing an international registry to reduce kidney trade at the international border could help reduce transplant tourism. Similar recommendations were also made by Martin *et al.*<sup>29</sup>

We confirmed that an overwhelming number of reported kidney sales involved Pakistani sellers and Middle Eastern buyers, which has been documented in the literature.<sup>30</sup> This trend may reflect the close relationship between



**Figure 5** Kidney trade networks using data from Region II news articles.

Pakistan and the Middle East in their religious beliefs. This may also reflect the fact that many Pakistani migrant workers live and work in the Middle East, through which the countries have built strong bilateral relationships over time. It may also be worth noting that a large portion of medical doctors and staff in the Middle East are of Pakistani origins<sup>31 32</sup> who can potentially connect the kidney patients from the region to a potential seller in Pakistan. Most of the surgeries occurring in Pakistan for Middle Eastern buyers may reflect the challenge/complexity that Pakistani sellers could face in obtaining visa to travel to the Middle East. Moreover, the cost of kidney transplantation is also considerably cheaper in Pakistan than in the recipient's origin country.<sup>33 34</sup>

We note several limitations in the data and the approach, as summarised below:

1. Incidences reported in newspapers do not provide a comprehensive picture of kidney trade/trafficking networks. Incidences in specific countries or regions are less likely to be revealed or reported due to corruption or the level of law enforcement. Such biases are, however, inevitable in any data on clandestine operations. In general, information sources on these activities are limited to: (1) court materials or official reports on revealed cases; (2) news articles; and (3) academic case studies. Each of these sources has advantages and disadvantages: while court materials could provide more reliable and in-depth information on each case, they are often unavailable for various reasons such as the case is in process/suspended, or there is no legal mandate to make such materials accessible to the general public. Academic case studies involving interviews and/or surveys could also provide detailed information than news articles. However, these cases are often selected for study among previously reported cases including those appeared in news articles. As such, it is highly unlikely that web scraping of academic publications that report specific kidney trafficking cases could provide a more comprehensive picture of kidney trade networks. Furthermore, court materials, official reports and academic publications have inherent time lags between event and reporting times when compared with news articles. One potential future direction is to web scrape all these information sources to compile more comprehensive and balanced records of kidney trafficking incidences. Such effort, however, will need to accompany a systematic way to eliminate duplicate records from the combined data.
2. The selection of newspapers can be an initial source of bias as some newspapers are more likely to report kidney trade cases than other newspapers, presumably due to the editors or reporters who are more familiar with or interested in the topic. Careful selection of the newspapers involving local experts should ameliorate the bias to some extent. The selection of only English language newspapers may also generate biases since newspapers in local languages could report a higher number of trafficking events due to their focus on regional affairs. As mentioned, the current study is exploratory in nature and thus focused on English language newspapers based on the recommendation of local experts who stated that the contents of English and local language newspapers overlap significantly. Following such a recommendation was deemed appropriate as we hoped to minimise duplicates in our records. A future full-scale study should, however, involve a thorough assessment on the degree of the content overlaps between English and local language newspapers along with the effort to include relevant local language newspapers in collaboration with local language speakers. In addition, exclusion of regional newspapers in Indian states of Punjab, Rajasthan, Gujarat, and Jammu and Kashmir may have influenced our findings for the Region 2 networks. Future endeavour could include newspapers from these Indian states to improve the understanding of South Asian kidney trade hubs.
3. It was inevitable that certain cases were counted more than once. We learnt that some notorious cases of kidney trade are reported multiple times by multiple newspapers. These cases also appeared in the articles that referred to historical cases of kidney trafficking. While this bias would not affect the list of countries involved in each network, the number of times that each country appears in the networks as a seller, buyer or surgery country, as well as the number of ties between countries, may have been affected.
4. Online censoring is another source of bias. The cases of those countries implementing extensive censoring are much less reported in the news, and thus the cases of those countries are erroneously under-represented in our data. Of note, all article collections under the current study were done in the USA, and thus the bias due to online censoring should be less than the scenario where collections were performed locally. That being said, as some countries in our data perform more censoring than others, any potential biases attributable to online censoring need to be carefully reflected on when interpreting our results.
5. Finally, the kidney trade hot spots emerge and disappear dynamically over time depending on various conditions of the countries. The kidney trade networks presented here provide only a snapshot of the networks. Thus, if the findings were to be used for law enforcement, the analysis with up-to-date data would be warranted.

Despite these caveats, our study demonstrated that web scraping data from news articles are a promising approach to collect data on kidney trade networks and to study characteristics of human kidney market and its network. Web scraping has been used extensively in many areas of research including labour market in the USA,<sup>35</sup> digital marketing,<sup>36</sup> rental housing<sup>37</sup> and crimes.<sup>38 39</sup>

In the field of other trafficking, Li *et al*<sup>40</sup> studied human trafficking networks operating on the adult service website using unsupervised scalable text template matching. The

study showed that the networks identified through the algorithm match well with surrogate truth data derived from phone number networks in the same corpus.<sup>40</sup> Similar work by Keskin *et al*<sup>41</sup> searched 10 million online classified sexual service advertisements to delineate the sex trafficking networks across US cities. Most, if not all, of the web scraping research in the trafficking arena has so far focused on the development of a tool to detect and eradicate specific illicit networks operating online. In contrast, the present study mainly focused on capturing the macropattern of kidney trade networks to gain new knowledge about the characteristics of the counties involved in kidney trade.

The present study has several directions for extension. Although the data presented here are limited in South Asia, analytical tools developed here can be applied to examine the structure and attributes of transnational kidney trade networks worldwide. Social network analysis could be applied to examine the structure of transnational kidney trade networks, as applied in a human trafficking setting by Goist *et al*.<sup>42</sup> In addition to social network analysis, various types of logistic regressions could also be applied to understand the determinants of a country becoming a seller, buyer or surgery country.

Several directions are possible to expand the data coverage. As mentioned above, expanded regional coverage and integration of data from court materials, official reports and academic journal publications as well as coverage of local language newspapers are some of the immediate next steps. Although the current study focused on collecting the information of seller, buyer and surgery countries only, the effort to include the nationalities of brokers as well as the countries of origin of involved crime organisations (eg, Ukrainian crime network involved in the Costa Rican case) can be made. Similarly, recording additional statistics, for example, which sections in the newspapers kidney trade articles appear, would also provide valuable insights into the ways to improve future data collections. Lastly, expanding the time coverage of the news article collection and recording the year of each reported case would allow us to examine the evolution of kidney trade networks.

## CONCLUSIONS

The findings of the present study suggest that newspaper scraping combined with the ML approach is a promising approach to compile data on illicit kidney trade, especially in the shortage of empirical data. While the present study was exploratory, further efforts along this line could provide a solid foundation for quantitative analyses of kidney trade networks. The data generated as a result of such efforts could help researchers examine the pattern of transnational kidney trade networks. We hope that the approach demonstrated under the current study will ultimately be helpful for eradicating illegal transplant tourism, but will also not be used to encourage travels for transplantation to non-studied regions.

The present study is the first attempt to delineate transnational kidney trade networks. We observed several factors and conditions affecting any given country to become a seller, buyer, surgery or broker country. The likelihood of a country to become a major surgery location seems to be driven by the county: (1) having ad hoc law enforcement; (2) having a good medical infrastructure; (3) being adjacent to poor countries; and (4) having established medical tourism business. Although these observations from the extracted networks warrant validations through further quantitative analyses, we opine that the findings together with the methodological approach presented here offer an important first step towards more systematic understanding on transnational organ trade networks and transplant tourism in a global context.

**Contributors** M-HL: conceptualisation, data cleaning and analysis, and writing. ABS: conceptualisation, data collection, visualisation and writing. BW and AP: conceptualisation and data collection. HE-A: conceptualisation and methodology. NK: conceptualisation, data collection and analysis, supervision and writing. NK is the guarantor of the paper.

**Funding** The study was funded/supported by grant 'Network Analysis and Opportunities for Disruption of Organ Trafficking' (EAGER: ISN:1838306) from the National Science Foundation, USA.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. Data and codes are available by emailing [mli11@gmu.edu](mailto:mli11@gmu.edu).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Abu Bakkar Siddique <http://orcid.org/0000-0002-9964-7511>

Naoru Koizumi <http://orcid.org/0000-0001-8722-0898>

## REFERENCES

- 1 Organization WH. *Resolution on human organ and tissue transplantation*. 17. Geneva, 2014.
- 2 Danovitch GM, Chapman J, Capron AM, *et al*. Organ trafficking and transplant tourism: the role of global professional ethical standards—the 2008 Declaration of Istanbul. *Transplantation* 2013;95:1306–12.
- 3 Thomson AM. A Summit on organ trafficking and transplant tourism: the Vatican Declaration on organ transplant. *Glob Kidney Exch Anal Backgr Pap Perspect Med Anthropol* 2017;70.
- 4 Columb S. Excavating the organ trade: an empirical study of organ trading networks in Cairo, Egypt. *Br J Criminol* 2017;57.
- 5 Sanchez R. United Nations investigates claim of ISIS organ theft. CNN [Internet]. 2015 [cited 2020 Sep 19]. Available: <https://www.cnn.com/2015/09/18/health/isis-organ-theft/index.html>.



- cnn.com/2015/02/18/middleeast/isis-organ-harvesting-claim/index.html
- 6 Fraser C, Koizumi N. Human organ trafficking: a study of change in the Egyptian kidney market 2008-2016. *Transplantation* 2017;101:S13.
  - 7 Cook M. BioEdge: Costa Rica becoming hub for global organ trafficking. BioEdge [Internet]. 2017 Sep 30 [cited 2020 Sep 19]. Available: <https://www.bioedge.org/bioethics/costa-rica-becoming-hub-for-global-organ-trafficking/12451>
  - 8 Moniruzzaman M. "Living cadavers" in Bangladesh: bioviolence in the human organ bazaar. *Med Anthropol Q* 2012;26:69–91.
  - 9 Hott Project. Combating trafficking in persons for the purpose of organ removal. In The Hague. The Netherlands; 2014 [cited 2021 Mar 9]. Available: <http://hottproject.com/reports/reports.html>
  - 10 Moniruzzaman M. At the Organ Bazaar of Bangladesh: In Search of Kidney Sellers. In: Siegel D, de Wildt R, eds. *Ethical Concerns in Research on Human Trafficking [Internet] (Studies of Organized Crime)*. Cham: Springer International Publishing, 2016: 227–47. [https://doi.org/10.1007/978-3-319-21521-1\\_14](https://doi.org/10.1007/978-3-319-21521-1_14)
  - 11 Moniruzzaman M. The trade in human liver lobes: bioviolence against organ sellers in Bangladesh. *J R Anthropol Inst* 2019;25:566–86.
  - 12 O'Connor N. Bangladesh organ trade continues unabated, targeting children, the poor. Aljazeera America [Internet]. 2014 Dec 26 [cited 2022 Apr 8]. Available: <http://america.aljazeera.com/articles/2014/12/26/bangladesh-organtrafficking.html>
  - 13 Srinivasan S. HEALTH-SOUTH ASIA: Hub For Global Organ Trade | Inter Press Service. Inter Press Service [Internet]. 2008 Feb 20 [cited 2020 Sep 20]. Available: <http://www.ipsnews.net/2008/02/health-south-asia-hub-for-global-organ-trade/>
  - 14 Adhikari B. Organ and human trafficking in Nepal. *Lancet* 2016;387:1907.
  - 15 Shrestha B, Adhikari B, Shrestha M, et al. Kidney sellers from a village in Nepal: protocol for an ethnographic study. *JMIR Res Protoc* 2022;11:e29364.
  - 16 Flaherty GT, Nasir N, Gormley CM, et al. Transplant tourism and organ trafficking: current practices, controversies and solutions. *Int J Travel Med Glob Health* 2021;9:102–6.
  - 17 Azad MAK. Bangladesh [Internet]. Media Landscapes. [cited 2022 Jul 11]. Available: <https://medialandscapes.org/country/bangladesh/media/print>
  - 18 Genilo JW, Asiuzzaman M, Osmani MMH. Small circulation, big impact: English language newspaper readability in Bangladesh. *AJC* 2016;04:127–48.
  - 19 Daily-Sun. Kidney trafficking goes unchecked | Daily Sun. daily-sun [Internet]. 22nd November, 2021. 2021 [cited 2021 Dec 4]. Available: <https://www.daily-sun.com/post/589485/Kidney-trafficking-goes-unchecked>
  - 20 Lewis DD, Yang Y, Rose TG. Rcv1: a new benchmark collection for text categorization research. *J Mach Learn Res* 2004;5:361–97.
  - 21 Porter MF. *Snowball: a language for stemming algorithms*, 2001.
  - 22 Bonaccorso G. *Machine learning algorithms*. Packt Publishing Ltd, 2017.
  - 23 Kuhn M. Caret: classification and regression training. *Astrophys Source Code Libr* 2015:ascl-1505.
  - 24 Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol Modell* 2003;160:249–64.
  - 25 Goyal M, Mehta RL, Schneiderman LJ, et al. Economic and health consequences of selling a kidney in India. *JAMA* 2002;288:1589.
  - 26 Budiani-Saberi DA, Delmonico FL. Organ trafficking and transplant tourism: a commentary on the global realities. *Am J Transplant* 2008;8:925–9.
  - 27 Cohen IG. Transplant tourism: the ethics and regulation of international markets for organs. *J Law Med Ethics* 2013;41:269–85.
  - 28 UNODC. Multi Country Study of the Smuggling of Migrants and Trafficking in Persons from Nepal. [Internet]. Regional Office for South Asia.; 2019 [cited 2022 Apr 8]. Available: [https://www.unodc.org/documents/southasia//Multicountry\\_study\\_TIP\\_SOM\\_NEPAL\\_2019.pdf](https://www.unodc.org/documents/southasia//Multicountry_study_TIP_SOM_NEPAL_2019.pdf)
  - 29 Martin DE, Van Assche K, Dominguez-Gil B, et al. Prevention of transnational Transplant-Related Crimes-What more can be done? *Transplantation* 2016;100:1776–84.
  - 30 Naqvi SAA, Ali B, Mazhar F, et al. A socioeconomic survey of kidney vendors in Pakistan. *Transpl Int* 2007;20:934–9.
  - 31 Chaudry S. Hundreds of Pakistani doctors brace for sudden deportation from Saudi Arabia. Middle East Eye [Internet]. 2019 Sep 1 [cited 2021 Mar 12]. Available: <http://www.middleeasteye.net/news/pakistani-doctors-brace-uncertain-future-wake-saudi-decision-reject-their-medical-degrees>
  - 32 Gilani I. Axe won't fall on Pakistani doctors working in KSA, Middle East. The Nation [Internet]. 2019 Aug 17 [cited 2021 Mar 12]. Available: <https://nation.com.pk/17-Aug-2019/axe-won-t-fall-on-pakistani-doctors-working-in-ksa-middle-east>
  - 33 Siddiqui N. UAE suspends issuance of visit visas to 12 countries including Pakistan, says FO. DAWNCOM [Internet]. 2020 Nov 18 [cited 2021 Mar 12]. Available: <https://www.dawn.com/news/1591082>
  - 34 The Associated Press. UAE Top Diplomat Acknowledges Visa Restrictions on Pakistan. Bloomberg.com [Internet]. 2020 Dec 21 [cited 2021 Mar 12]. Available: <https://www.bloomberg.com/news/articles/2020-12-21/uae-top-diplomat-acknowledges-visa-restrictions-on-pakistan>
  - 35 Atalay E, Phongthientham P, Sotelo S, et al. The evolution of work in the United States. *Am Econ J Appl Econ* 2020;12:1–34.
  - 36 Vargiu E, Urru M. Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artif Intell Res* 2013;2:44–54.
  - 37 Boeing G, Waddell P. New insights into Rental housing markets across the United States: web Scraping and analyzing Craigslist Rental Listings. *J Plan Educ Res* 2017;37:457–76.
  - 38 Chakravorty S, Daripa S, Saha U. Data mining techniques for analyzing murder related structured and unstructured data. *Am J Adv Comput* 2015;2:47–54.
  - 39 Rahman RU, Tomar DS. A new web forensic framework for BOT crime investigation. *Forensic Sci Int* 2020;33:300943.
  - 40 Li L, Simek O, Lai A. Detection and characterization of human trafficking networks using unsupervised scalable text template matching. In: *2018 IEEE International Conference on big data (big data)*. IEEE, 2018: 3111–20.
  - 41 Keskin BB, Bott GJ, Freeman NK. Cracking sex trafficking: data analysis, pattern recognition, and path prediction. *Prod Oper Manag* 2020.
  - 42 Goist M, Chen THY, Boylan C. Reconstructing and analyzing the transnational human trafficking network. In: *2019 IEEE/ACM International Conference on advances in social networks analysis and mining (ASONAM)*. IEEE, 2019: 493–500.