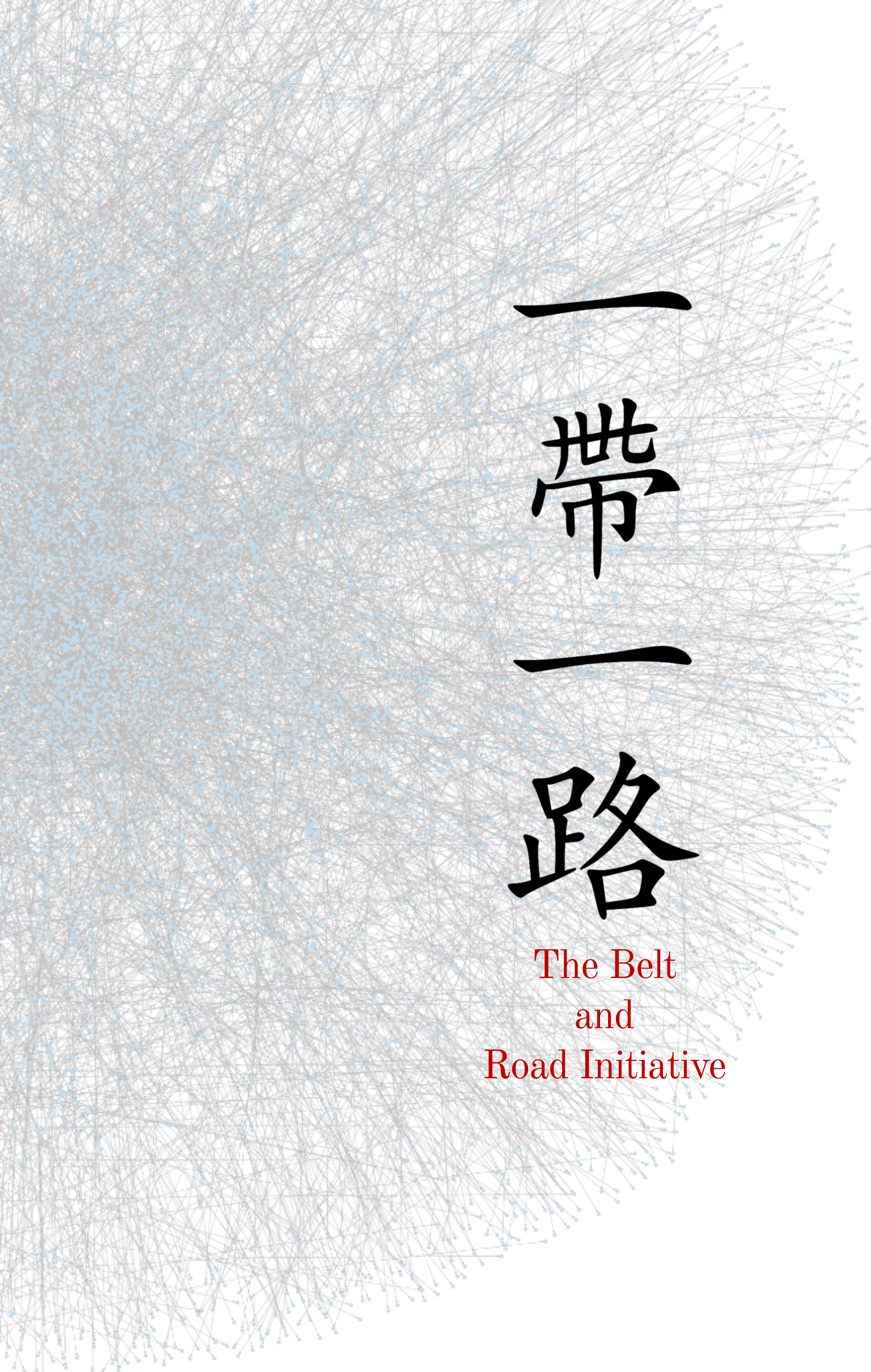


一帶一路

The Belt and Road Initiative

Towards a systems-based understanding of international trade

The role of infrastructure and distance in the case of the Belt and Road Initiative

This dissertation is submitted in part requirement for the MSc Spatial Data Science and Visualisation in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

CASA0012 | 24.08.2020

Antonios Fiala 18131325 | Supervised by Dr Thomas P. Oléron-Evans | Word count 11,958

Učit se, učit se, učit se
教育改变命运

CASA0012 | 24.08.2020
Antonios Fiala 18131325 | Supervised by Dr Thomas Oléron Evans | Word count 11,958

Towards a systems-based understanding of international trade

Examining the role of infrastructure and distance in the case of the Belt and Road Initiative

This paper address calls in literature to incorporate transport literature into studies of international trade and extends this by advocating for a systems-based solution. The study focuses on the effects of different measures in international trade literature on the collection of trading entities as a whole and specifically in identifying most efficient paths across which goods could be transported. To achieve this, a proof of concept network is created based on the Belt and Road Initiative at country and route level and outcomes compared.

Findings suggests that route-level analysis could potentially lead to more accurate estimations of countries along routes as well as the distance travelled.

Finally, as a stepping stone towards a means of examining international trade from a systems point of view, the framework and data processing flows are proposed as a baseline on which future complementary work can expand based on the published material on GitHub.

I hereby declare that this dissertation is all my own original work with all sources acknowledged.
It is 11,958 words in length.

Accompanying GitHub repository: https://github.com/antoniosfiala/UCL_Dissertation

Content

Tables |

Figures |

Page intentionally left blank

1 | Introduction

In today's globalised world economy and changing global power dynamics the role of international trade in ensuring economic security, productivity and growth is vital and is a field of study that maintains its proportionate share of academic attention. However, although traditional methods such as gravity modelling are growing in sophistication, they appear to be confined to country-level analyses. With globe-spanning initiatives such as the Belt and Road Initiative (BRI), spearheaded by China since 2013, coming to the fore, it is essential to bridge disciplinary silos and construct a means of modelling and understanding the impact of such projects on trade and surrounding systems.

This paper contributes to this research gap in several ways. Firstly, by bringing together essential concepts across a number of disciplines. These are discussed in detail throughout Section 2 and culminate in the placement of the BRI into this context. They cover from geopolitical considerations that help us understand *why* the BRI is significant but also *why* it could take place only in the last decade. Further exploration draws on international trade literature and builds on the agreed-upon consensus of crucial determinants of trade to then scope out the relevant parts of the BRI for this study which are infrastructure connectivity and trade facilitation. The relevance of the concepts at the start while not direct to the end findings is crucial in providing a roadmap for extendibility and data set considerations in the future.

Finally, this paper emphasis reproducibility and aims to not only clearly outline its methodology but data processing so that it can be further developed and extended, all accompanying work files can be found in the marked GitHub repository.[1]

The specific questions addressed revolve around the role of distance and infrastructure in the study of international trade which is explored through network analysis. With a geographical scope of Europe and Asia to capture the central actors of the BRI, the emphasis is placed on constructing a proof of concept model through the utilisation of country-level datasets and route-level (rail infrastructure) datasets with the aim of conducting a comparative analysis of their respective results. While findings are encouraging in the added insight to be gained from route-level analysis the limitation currently lies in the connections considered and future work ought to extend these to other land, air and sea routes through which trade could flow.

[1] GitHub repository: https://github.com/antoniosfiala/UCL_Dissertation

2 | Literature review

2.1 | Overview

What does this section cover?

The Belt and Road Initiative (BRI), although an infrastructure and trade initiative at its core, is an endeavour of evolving scope, magnitude and impact across political, economic, cultural and environmental spheres to name a few. The constructive ambiguity around the BRI's evolving nature has captured the imagination of policymakers, analysts and journalists alike from its inception and continues to frame contemporary great-power analysis to this day (Bokhari et al., 2020; Foroohar, 2020; Griggs, 2020; Hille et al., 2020). While this prominence has led to expected academic attention, many have - understandably - focused on their specific domains and pointed research questions, for instance:

- What are key *environmental* concerns across a region X? (Hughes et al., 2020)
- What are the *geopolitical* implications of BRI? (Anastasiadou, 2019)
- What are the *economic* benefits for country Y as a result of increasing *trade*? or Who are the *material* beneficiaries? (Lall and Lebrand, 2019; Ruta et al., 2019)

However, to understand the BRI, or any other comparable construct, in its entirety, it is proposed that a systems approach ought to be used which can gradually build up the various interacting dimensions of a complex and multifaceted space. This section brings together the necessary concepts across disciplines to provide a broad contextual background and firm grounding on which to unravel the BRI and present research questions around the core of the BRI, infrastructure and trade, with potential implications on the study of international trade as a whole.

This conceptualisation is inspired by the work Lippe et al. (2019) who introduce Social-Ecological Systems (SES) in an attempt of modelling "human and environmental systems that mutually influence each other." Their definition continues as per the extract below (underline added):

> *"An SES in this view includes the ecological components of an interdependent group of organisms or biological entities, within a bio-geophysical environment [5, 6]; and a social component including the actors whose activities directly influence ecosystems and those that govern human-nature interactions which can be the same or different actors. Resulting interactions are mediated by the broader social, economic, and political settings and the larger ecosystems within which the SES is embedded [7]. <u>Interactions are continuously changing due to feedbacks and internal or external factors, taking place across different temporal and spatial scales, making SESs highly dynamic systems</u> [8–10]." (Lippe et al., 2019)*

In this case, components that make up the system on hand and interact are different, although to an extent they open up the socio-economic parts of an SES. The macro components are (i) geo-politics and (ii) international trade. In line with the above, these are systems where the interaction is continuous and take place across temporal and spatial scales thus also making them highly dynamic. The final component is (iii) the BRI itself which is a system which overlaps several sub-systems of component (i) and (ii). Figure 1 demonstrates this below.

Figure 1 |
View of the relevant systems, literature and the focus of this work

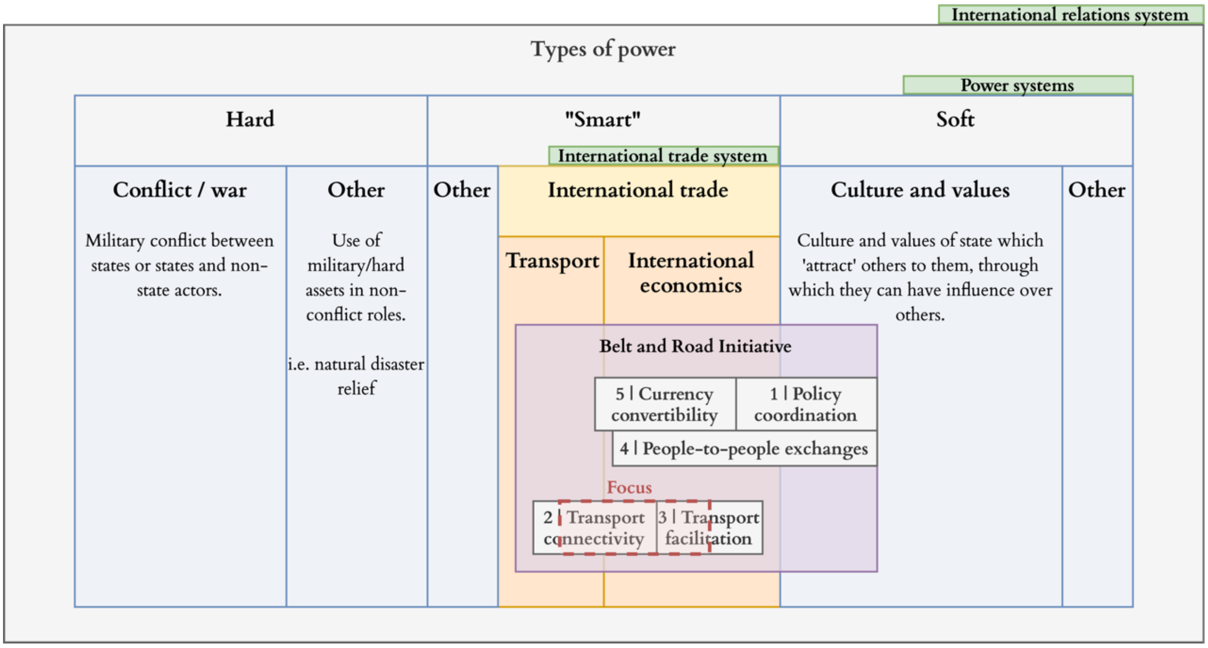


Figure 1 highlights the hierarchy of the macro systems from geo-politics and typologies of power down to international trade and the elements (1,2,3,4,5) that are at the heart of the BRI and the focus which is around transport connectivity and facilitation. As mentioned above, given the many moving parts and concepts, the literature review will introduce the key concepts of each system component (i, ii, iii) in turn in order to sufficiently cover the broader picture and illustrate the relevance of each block deeper down.

Finally, it must be acknowledged that it is beyond the scope of this work to recreate a system that encapsulates all the components outlined in Figure 1. That is a limitation that can only be addressed through further work.

2.2 | Power, key developments and systems theory in geo-politics

What geo-political concepts help us understand BRI? What role has systems theory played to date?

The disciplines of international relations and geo-politics are, at their core, the study of the distribution of power in the global system and how the primary international actors behave within them. Traditionally, the *state* has been both the central actor of analysis and means of organising global phenomena (Giles, 1970). The objectives of states can be achieved through an array of strategies that are generally presented across a spectrum of power ranging from 'hard' to 'smart' and 'soft' as illustrated in Figure 1 above (Baldwin and Kay, 1975, p. 30).

Traditionally, power was studied as a dichotomy of hard (military) or soft (values) with either beings means to particular ends. However, the end of the Cold War brought with it a broader (globalisation-driven) structural change in the landscape of international relations that cleared the way for a third, so-called 'smart', way of exerting power (Nye Jr, 2009). Nye's 'smart power' revolves around the premise that an actor can either provide, or coordinate action to secure so-called 'public goods' that no one actor could individually secure. In the economic definition, these are goods that can be consumed in a non-rival fashion and are (typically) non-excludable due to which the incentives for any single actor to provide them are low (Kaul et al., 1999).

Table 1 explains these key changes that have made allowed China to rise and secure its place as a key international power over the past 30 years.

Table 1 |
Changes in global power relations

Change	Explanation
Centrality of economic power *Move away from emphasising hard (military) power as primary means of deterrent and achieving key objectives.*	The increasing levels of globalisation and country interdependence have led to a shift in emphasis of type of power in geo-politics. While military means are prevalent, the associated destruction and disruption has led to a shift, particularly with newer actors (such as the European Union) or newly risen powers (BRICS), to economic means of exerting influence and achieving strategic objectives (Meunier and Nicolaïdis, 2005; Nye Jr, 2009; Stavridis, 2001). This speaks more to the emphasis of a particular toolkit in contrast to the past rather than the absence of other types of power from contemporary geopolitics.
Entry of new powers *Brazil, Russia, India, China, South Africa (BRICS).*	With the increase of prominence of economic power, this has allowed other actors, most notably the BRICS countries, to begin exerting additional influence in their neighbourhoods, giving rise to new 'resource rich' regional powers (Wilson, 2016). The extent of their power however continues to be debated and suggested the monolithic grouping of these states may not be appropriate (Glosny, 2010; Jacobs and Van Rossem, 2014).
Multipolarity and rebalancing *The redistribution of power from two, to one, to many(?) actors.*	The combination of the above changes has resulted in a reshaping of the global system. Moving beyond the bi-polar power arrangement of the Cold War, the moment of unipolarity of the US continues, albeit within a context of rebalancing/redistribution to most notably China as the foremost BRICS country (Glosny, 2010). The consequence of this new reality is that the actions of stronger actors can now be felt. Returning to systems theory, this gives their actions greater weight and can therefore lead to shaping of the system they inhabit form within as well as potentially new system outputs. Arguably, the BRI as a sub-system of international trade, is a means through which China's smart power could be amplified beyond the aggregate power of the BRI members.

In an environment defined by multipolarity and the centrality of economic power, China's decision to pursue the BRI marked a departure point from a decades old foreign-policy doctrine of 'biding its time' and low profile (as espoused by Deng Xiaoping) and signalled the beginning of major interventions (Harshaw, 2018). The BRI can be seen as a means of pursuing a Chinese agenda while mobilising actors to generate and invest in public goods such as infrastructure.

The assertion of this policy from China is only possible due to the different nature of power within the geopolitical system. Despite the clear system qualities of state-to-state interactions, recent literature such as Jervis (2010) and Harkavy (2013), while speaking to the utility of systems approaches in geo-politics, concludes in highlighting its ongoing underutilisation. An examination of past literature, suggest this to be a long-present research gap (Flint, 2001; Starr, 1991).

This reflection touches on the nature of power in geo-politics and the centrality of 'smart' (economy-centred) power. In this climate the move of China to pursue an initiative like the BRI speaks to its desire to perpetuate and shape the system that enabled to re-emerge as a key international actor. Further exploration of the literature shows an underutilisation of systems approaches, particularly those that would connect the underlying economic mechanisms. It is to economics and international trade that the following section turns to in order to understand its key concepts but key factors that the BRI speaks and seeks to shape.

2.3 | Systems and the key concepts of international trade

How has complexity been dealt with in economics? What drives international trade?

The state of economics with regards to systems theory is comparable to that of geopolitics in Section 2.2. Calls go as far back as the 1970s and 1980s for the integration of the two to address the "underdeveloped" nature of systems in economics (Hodgson, 1987; Kornai, 1971). These calls appear to be ongoing with recent works of Foster (2005) and Valentinov (2014) corroborating the state of the discipline.

The work of Foster (2005) in particular is important to dwell on as they clearly compare and contrast the optimisation driven economic schools of thought with system-centred economic schools of thought, providing definition and examples along the way. Crucially, it is argued that both have a role to play with the former being particularly suited for short-run and local economic deployment given the smaller chance of change in the fundamental nature of the system (Foster, 2005, p. 811). In establishing a steady theoretical foothold, it is both important to recognise the existence, although not necessarily widespread, of systems theory with an economics twist (that is different from traditional systems rooted in physical, chemical or biological environments) and the place traditional optimisation models play a role. Table 2 alongside the commentary below outlines the fundamentals of international trade through which the applicability of systems and in particular the centrality of *networked connections* can be deduced. [2]

Looking at drivers of trade, the theoretical foundation revolves around the concept of 'comparative advantage' that makes trade worthwhile(Hausmann et al., 2019). The go-to methodology using gravity equations has reduced the determinants of trade to the size of each economy, the distance between origin and destinations, and other factors broadly labelled as 'trade frictions' shown to be able to broadly account for changes in trade (Baier et al., 2018). Of note is the ongoing debate around 'distance' as quite often the utility of point-to-point geographic distance does not always account for a full picture which as a result continues to attract research (Rodrigue, 2012). Finally, there is a strand of literature which critiques the traditional study of international trade that omits the explicit inclusion of transportation literature findings as argued by Baier et al. (2018).

[2] More suggested literature on international trade includes works by Blonigen and Wilson (2018), Martin (2015) and Van Bergeijk and Brakman (2010)
For further reading on systems theory in economics, the works referenced in Foster (2005) provide a useful overview of key texts

**Table 2 |
Key international trade concepts**

Concept	Explanation
Comparative advantage *Why nations trade.*	A concept at the heart of international trade first introduced by David Ricardo circa 1817 which helps explains why countries trade based on the varying value of commodities, resource endowments and capital (altogether yielding relative productivity) across countries (Isgut et al., 2005; Ricardo, 1891). Comparative advantage is a natural extension of the principle of 'division of labour' at an international scale (Smith, 1776). Ricardo argues that specialisation in the productions of goods where a country has a relative productivity advantage (can produce a given commodity more efficiently) will allow it to maximise benefits through trade (Hausmann et al., 2019). Notably, a large scale multi-country and multi-commodity model has only been produced in the last decade (in contrast to the theory introduced over 200 years ago) (Hausmann et al., 2019). A closely related strand of research is understanding the causal mechanisms by which the disparities emerge in the first place. These are typically found to span a 'geography versus institutions' spectrum that each explain varying endowments of technological capital that ultimately serves as a principal explanatory variable for the disparities between nations (Costinot, 2009).[3]
Gravity modelling *Key factors determining trade.*	A model widely accepted to be "the workhorse" of empirical international economics for more than five decades (Baier et al., 2018, p. 15). Based on Newtonian physics and the law of attraction, it states that: *"Bilateral exports are proportional to economic size and inversely proportional to geographic distance."* (Chaney, 2018) Typical findings suggest that the most prominent role in determining trade can be assigned to three factors: (i) Size (or relative size) of each economy in a bilateral trading pair (ii) Distance (iii) Other variables denoting trade frictions Factor (iii) can include but is not limited to infrastructure, geographical/physical barriers, political barriers, differences in standards, administrative procedures, language which are also areas of discussion in areas such as supply chain (Lee and Swagel, 1997; Rodrigue, 2012).
Distance *A "mysterious" component.*	A factor shown (as per above) to explain the volume of trade. However, despite a traditional use of geographical distance between specific points, populated weighted points or capital cities (Berthelon and Freund, 2008; Yotov, 2012), there is a nuanced debate and work on refining to directly account for either route-level means of understanding distance such as Rodrigue (2012) from a supply chain perspective as well as international trade research such as Chaney (2018) outlining conditions when the effect of distance may be constant or inversely proportional as per the gravity definition above.
International trade and transport *How nations trade.*	The means through which nations trade, including roads, rail networks, as well as custom checks. The means of connecting countries through transportation links and the study of transporting goods along them is not always included, despite acting as an important component to trade friction (iii) in the studies of international trade, resulting in potential omissions as argued by Baier et al. (2018). This is an important lens through which to assess further works and assess potential research gaps on the BRI.

The implication of the above overview combined with the detail in Table 2 is that the definition of distance is inconsistent. In contrast to arguments dismissing the role of distance in a globalised world economy (Hess et al., 2013), the work of Borchert and Yotov (2017) and Magermann et al. (2016) put forward the view of distance varying by country and potentially industry as well. Furthermore, the arguments of Baier et al. (2018) are particularly relevant for the BRI as they logically lead one to not centre their investigation around monolithically-perceived state entities (as the case may be in both international trade and geo-political arenas) and instead, urge one to conduct network centred analysis that aim to model the system as a whole. This is precisely the focus of this work, focusing on the role of different types of distance across spatial scales and the commonly missing incorporation of transport links.

With the high-level theoretical framework in place, the subsequent section will outline the BRI itself and home in on the research gap observed and resulting research questions.

[3] On the geography versus institutions debate, following provide a grounding: Diamond (2005), Kourtellos et al., (2010), Robinson and Acemoglu (2012) and Rodrik (2004).

2.4 | Understanding the fundamentals of the Belt and Road Initiative

What is the Belt and Road Initiative? Why does it matter?

The initiative for building *The Silk Road Economic Belt and the 21st Century Maritime Silk Road* was unveiled in 2013 by Chinese President, Xi Jinping in two stages. An initial state visit to Kazakhstan which emphasised the importance of strengthening overland connections across Eurasia (the belt) and a subsequent visit to Indonesia introduce the concept of a Maritime Silk Road with a clear and strong emphasis on infrastructure construction. Rebranded three years later to *The Belt and Road Initiative* (BRI), its scope and ambition continue to be loosely defined.[4] As a consequence of this constructive ambiguity, the BRI has been expanded in terms of both scope and geographical coverage continuing to fuel debates around its true motives and potential implications on the state of the global economy (Frankopan, 2019, 2015; Sterling, 2018; Wang, 2020; Zeng, 2019). The dynamic nature of the BRI alone would warrant academic focus, however given its scale within a system of geopolitical and trade arrangements, the question is open as to whether there can be system-altering as opposed to rebalancing repercussions.

Whether in its 2013 state or the 2020 iteration, the word 'connectivity' can be said to adequately summarise the objective of the BRI (Gabusi, 2017). Figure 2 outlines the key sated components of the BRI, highlighting those most relevant to this work highlighted in red with Table 4 going into more detail (Huang, 2016).

Figure 2 |
Five Belt and Road Initiative components (五通)

Source: Hughes et al. (2020), visual representation by this author.
Note: The components in red are the main focus of this work.

Seeking to quantify elements of the BRI remains a challenge due to its ambiguity. However, a number of sources agree that the approximate scale of the BRI as well as China's fewer prerequisites when it comes to initially favourable financing arrangement are two key differentiators compared to other initiatives (Mobley, 2019). This is significant as it is an important factor in less economically countries agreeing their membership.

Table 3 |
The Belt and Road Initiative in numbers

Measure	2020 / latest figures
Countries	30 core countries (Brakman et al., 2019), 65 countries (Dumor and Yao, 2019), 138 (IIGF, 2020) based on signed count of Memoranda of Understanding[5]
Investment committed[6]	~ USD 1 trillion (EBRD, 2020; OECD, 2018; PwC's GWC, 2016; Silk Road Briefing, 2020)
Number of projects	2,951 based on tracking at Refinitiv and reported by Husein et al (2019) but difficult to corroborate across multiple sources.

[4] Naming conventions for the BRI are converging from the initial 2013 announcement of a land belt and maritime road, often referred to as 'One Belt, One Road' evolved to the now-more-commonly encountered Belt and Road Initiative (BRI) since guidance was issued by the Central Compilation and Translation Bureau of the People's Republic of China and the Chinese Academy of Social Sciences in 2016 (Bērziņa-Čerenkova, 2016; Stanzel, 2017)

[5] Source data, links and validation of data by the author available in GitHub repository as part of raw data. However, the signing of these documents does not necessarily suggest concrete steps are taken. Equivalent found in mandarin can be found on a government portal (Belt and Road Portal, 2019).

[6] Investment to date is not easy to track for two reasons. Firstly, it is difficult to separate any domestic or Foreign Direct Investment (FDI) which materialised solely due to the BRI. Secondly, the loose definition of the plan and risk with each project does not guarantee all the funds will be able to be deployed. Nevertheless, it is generally understood that the quantum of investment is vast and will go some way to address the infrastructure investment shortfall of the BRI countries.

Table 4 |
Belt and Road Initiative components expanded and explained

Priority areas	Sub-area	Focus
1 | Policy dialogue	**1.1 | Macro-policy Forum**	Forum for debate and coordination of collective action direction. Macro-policy dialogue, deepening of trust and shared interests to reach consensus and promote political trust.
	1.2 | Monitor/support large projects	Means of monitoring implementation of defined objective/project in 1.1
2 | Infrastructure connectivity	**2.1 | Harmonise technical standards**	Define common standards to increase efficiency and interoperability of components across the region.
	2.2 | Formation of routes/connections	Create new connections or upgrade existing connections across transport and energy links
3 | Unimpeded trade	**3.1 | Investment**	Focus on investment (easier flow of capital) to deploy it in ways that lower barriers to trade.
	3.2 | Lower barriers to trade	Administrative and infrastructural links such as border crossings
	3.3 | Free trade zones	Another means through which to lower trade friction
	3.4 | Other	Includes, promotion of integrated R&D, regional industrial compatibility Formation of cross-country industry value chains
4 | Financial support	**4.1 | Currency convertibility and settlement**	Currency convertibility and settlement; Asian bond market (including cross-border RMB denominated bonds); A development bank; Banking consortium;
	4.2 | Asia bond market	Including the issuance of RMB-denominated bonds
	4.3 | Banking	AIIB and BRICKS New Development bank and China-ASEAN Banking Consortium
5 | People-to-people exchange (enhancing mutual understanding and trust)	**Cultural, academic and talent exchanges**	For further detail please see work of Huang (2016)
	Training (unspecified)	
	Media cooperation	
	Youth and women's dialogues	
	Cross-country tourism promotion	
	Disease control	
	Joint research centres for laboratories	
	Political party and parliamentary exchanges	

Note: primarily drawn from the work of Huang (2016) and Hughes et al (2020)

The breakdown of the BRI in Table 4 illustrates the broad area of policy it covers and speaks to why it spanned across 'smart' and 'soft' power in the system and literature topology in Figure 1. Building on the previous section and the gap in the study of distance and transportation links, the (sub-)components of greatest relevance for this work are 2.2 and 2.3 which speak to routes (distance) and the barriers to trade (friction).

Although the "BRI remains a puzzle" that is not fully conceptualised, the above speaks to its significance through scale and briefly strives to understand the BRI through its focus on infrastructure and trade that are central to this work (Yuan, 2020). Lastly, while that speaks to the aspired *ends* of the BRI, Zeng (2019, 2017) speaks to the *means* which are notably local in the delivery of the BRI components. Therefore, given the heterogeneous nature of the BRI implementation, as well as the distributed impact across regions, it further reinforces the argument for a bottom up study of the BRI focused on routes and local dynamics (Lall and Lebrand, 2019; Zeng, 2017).

2.5 | Key strands of literature around the BRI

What are the prominent research areas to date and what challenges and gaps can be observed?

As illustrated above, the "amorphous" nature of the BRI is mirrored in the broad research topics pursued which often follow the five key areas but also tackle higher-order systemic questions, be they ecological, societal or geo-political (Anastasiadou, 2019; Buckley, 2020; Hughes et al., 2020; Kirby and Van der Wende, 2018; Lee et al., 2018). Examining works around research agendas, further trends and areas of interest are highlighted that speak to the unique challenges present which are addressed in Section 2.6.

At an international relations level, Yuan (2020) corroborates the hard to define nature of the BRI, but also its dynamism that has resulted in redefinition of key elements in the wake of 'push back' from participating countries. However, it remains telling that seven years since the BRI's announcement, questions remain regarding the "nature, process, scope and implications" (Yuan, 2020). Although the focus here is on discussions with regards to connectivity tied to trade, there are nonetheless considerations to take away which as illustrated through earlier parts of the literature overview ought to be taken into account. Firstly, this revolves around the call around an "open-minded conceptualisation" of BRI which is understood and assessed through a combination of theoretical lenses and non-western perspectives, an important consideration in system construction. Secondly, it highlights the methodological issue of "premature assumptions" that is typically present due to insufficient data which speaks to the need to build a system in which such assumptions can be replaced with defined dynamics as and when data becomes available.

In particular, the contribution of Lee et al. (2018) not only focuses on the trade/transportation aspect of BRI, it assesses the state of the literature with particular reference to Chinese-language works that could otherwise be overlooked in research such as this one. Furthermore, it conceptualises the BRI in relatively 'practical' terms along the specific transportation corridors and directly speaks to the concepts of *distance* and *trade friction* discussed with regards to international trade in Section 2.3. This is undoubtedly a function of the transportation matter at heart, nevertheless it starkly contrasts the afore-discussed work by Yuan (2020) or that of Hughes et al. (2020) that have broader themes not bound by particular routes/corridors.[7] Ultimately, the construction or improvement of respective corridors is a tangible *end* of the BRI which can be traced, data gathered on and approached pragmatically as all research agendas call for and will be adopted as a framework in this work as outlined in Figure 3.

Figure 3 |
Sample BRI research agendas and proposed directions

Yuan (2020)

International relations perspective

1 **Nature and scope of the BRI**
Significant in correctly identifying toolkits, methodologies and key parts if to be approached from a systems perspective.

2 **The BRI as an output of China's foreign relations**
What are the outputs and how they link to other parts of global systems?

3 **The BRI as processes since 2013**
Evolution and alteration over time.

4 **The BRI as an input to the world**
Channels through which systemic outputs are transmitted and their impact.

5 The comparative research on the BRI
Regionalist perspective on similarities to other initiatives such as the E.U., ASEAN, potential for lessons and why and what makes them different

Lee et al. (2018)

Transportation perspective

6 **Impact of CPEC corridor (Western China to Pakistan)**
Impact on direction of cargo flows (away from traditional routes in South East Asia due to lowering of geographical distance in the west) and resulting volumes of trade that affect capacity.

7 **Impact of corridors on regional transport systems**
Role of intermodal transport solutions (such as connecting rail and sea routes).

8 **Coordination mechanisms**
Minimise trade friction for flow of goods across transient countries.

9 **Infrastructure investment and capacity along the BRI network**
How it can handle increased volumes, particularly in light of potential Free Trade Agreements.

10 **Identification of new efficient routes**
Consequence of greater network capacity and new ports of call may open novel and efficient routes to countries inside BRI and outside, how do we find them, where may they be?

[7] Chan (2018) is a counter-example of broader historical and geo-political themes examined through a corridor centred framework. Likewise, Hughes et al. (2020) works along similar lines, considering broader ecological and societal issues which may be of interest. Lee and Shen (2020) provide a complimentary research agenda from a supply chain / logistics point of view.

Considering Figure 3 in the context of the build-up of the section thus far, research agendas 6 through 10 (from Lee et al) speak to the concepts of distance and trade frictions that are central in affecting trade patterns. In this case, agendas 6 and 10 in particular are selected as points of focus that are likely to function as inputs into a broader global system (agenda 4 from Yuan). Following the principle of Baier et al. (2018) of ensuring trade and transportation research are approached together, a selection of works from trade and transportation literature will be assess to get an overview of methodologies, adherence to the Baier principle and the level at which data has been conducted at (regional, country, route level).

Table 5 builds on the work of Lee et al. (2018), adding additional works to categorise papers and suggest that there is a notable absence of a number of elements gradually introduced during this section.

Table 5 |
Expanding BRI trade and transport centred literature

	Citation	BRI context	Country	Route / City	Transport	Methodology Most prevalent approach/methodology	Comment Topic / Problem addressed by the paper | Key finding of research | Relevance of the work to this project
1	(Dumor and Yao, 2019)	✓	✓	✗	✗	SIM / Gravity modelling	**Topic / Problem** | Trade volume estimation and comparison of gravity and ML **Finding** | Machine learning techniques more effective **Relevance** | Tools to use for estimating flow changes
2	(T. Wu et al., 2020)	✗	✓	✗	✗	SIM / Gravity modelling	**Topic / Problem** | Study effect of different measures of distance on exports **Finding** | Using measure of 'national distance' instead of geographical **Relevance** | Redefining distance that can be applied at route level analysis
3	(Kohl, 2019)	✓	✓	✗	✗	SIM / Gravity modelling	**Topic / Problem** | Welfare impact of trade-cost reducing measures (infrastructure) **Finding** | Asymmetric country benefits, BRI found better for growth vs TPP **Relevance** | Limitation of the distance variable used, scope of further research
4	(Huang et al., 2019)	✓	✓	✗	✗	SIM / Gravity modelling	**Topic / Problem** | Trade volume forecast to five of China's neighbours **Finding** | Robust explanatory variables, stable export growth expected until 2022 **Relevance** | Only uses geographical distance, to be extended with infrastructure
5	(Herrero and Xu, 2017)	✓	✓	✗	✗	SIM / Gravity modelling	**Topic / Problem** | Trade creation estimate in response to lower transport costs **Finding** | EU (particularly landlocked states) will see largest benefits **Relevance** | Showcases absence of route-level analysis
6	(Molnár et al., 2019)	✗	✓	✗	✗	Trade intensity indices / models	**Topic / Problem** | Index creation for measuring trade intensity for key BRI states **Finding** | Trade intensity for China, Central Asia and EU not affected by BRI **Relevance** | Variables used, little effect by BRI between 2013-2017
7	(Yu et al., 2020)	✓	✓	✗	✗	Index construction and difference-in-differences estimate	**Topic / Problem** | Assessing change in export potential **Finding** | Specific provinces in China **Relevance** | Illustration of research agenda, omission of specific projects/route data
8	(Wang et al., 2020)	✗	✓	✗	✓	Spatial autocorrelation	**Topic / Problem** | Impact of transport infrastructure on growth in BRI countries **Finding** | Positive impact with regionally varied spill over effects **Relevance** | Different types of distance considered (economic, geographic, cultural…)
9	(Z. Wu et al., 2020)	✗	✓	✗	✗	Difference in differences and propensity score matching	**Topic / Problem** | Assess participation in broader system of Global Value Chains **Finding** | BRI significantly promotes participation in Global Value Chains in year 2 **Relevance** | Illustration of methodology and research question
10	(De Soyres et al., 2018)	✗	✓	✓	✓	Network analysis	**Topic / Problem** | Transport (shipping) time reduction at route and country level **Finding** | Aggregate reduction ~ 1-3% with corridors seeing ~10% time and cost ↓ **Relevance** | Work that most comprehensively includes multi-level data
11	(Yang et al., 2018b)	✓	✗	✓	✓	Network analysis	**Topic / Problem** | Reconstruct shipping service network and find optimal networks **Finding** | New optimal routes identified for the expected state of BRI **Relevance** | The need to have quick access to these networks in order to test scenarios
12	(Wen et al., 2019)	✓	✗	✓	✓	Network and Route Utility Function	**Topic / Problem** | Route selection (overland vs sea corridors) focused on China **Finding** | Advantages found for land corridors over traditional ocean routes **Relevance** | Generation of utility function, variables and data used for network
13	(Melecky et al., 2019)	✗	✗	✓	✓	Econometric simulations	**Topic / Problem** | Framework for cost-benefit analysis of infrastructure projects **Finding** | Test on corridor across Pakistan shows promise **Relevance** | Criteria used for assessing projects and help account for externalities
14	(Jiang et al., 2018)	✗	✗	✓	✓	Binary logit model	**Topic / Problem** | Route selection problem along on BRI corridor (case of rail vs sea) **Finding** | Subsidies help keep rail competitive, development needed in west China **Relevance** | Key rail routes, cost structure and their development + role of subsidies
15	(Yang et al., 2018a)	✗	✗	✓	✗	Qualitative combined with MCDA	**Topic / Problem** | Route selection / competitiveness of BRI and non-BRI routes **Finding** | Developing criteria on which to comprehensively compare routes **Relevance** | View of methodology
16	(Sun et al., 2020)	✓	✗	✓	✗	SVM classification	**Topic / Problem** | Understanding urbanisation along the BRI (not as consequence of) **Finding** | Based on historical trends, BRI routes will be along key megacity areas **Relevance** | Point of consideration vis-à-vis interaction with cities
17	(Soner, 2020)	✓	✗	✗	✓	Semi structured interviews	**Topic / Problem** | Port (read key transport node) selection **Finding** | Development of framework for port selection **Relevance** | Visibility of criteria that could play role in a broader system

Notable is the absence of systems approaches. However, given their rare application in related fields of geopolitics and economics (see Sections 2.2 and 2.3) this is not entirely surprising. Nevertheless, it ought to be a point taken onboard. Furthermore, in line with international trade literature, significant portion of literature addresses points of trade flows with reference to countries (i.e. works 1-9 in Table 5), however it tends to omit specific route/transportation elements as per the observation Baier et al. (2018) made for the discipline. Of note is the awareness of the authors for this limitation and suggestions for inclusion in further work. In contrast, and not that surprisingly (given the revelation from Lee et al. (2018)) works focused on the logistics / supply chain dimension (i.e. 10-14 / 17 in Table 5) utilise network approaches to study changes of impacts of the BRI at a detailed level, but with no direct linkage to higher-order phenomena such as inter-country trade.

Overall, it is not surprising to find that a large body of works exists focusing on international trade or infrastructure. The scale and emphasis of these two domains by the BRI has attracted the expected academic attention. However, it is rare, for the work on BRI and more broadly to see trade and infrastructure work to be considered in tandem. The nascency of the BRI as a subject of study can go some way to explaining the more siloed approaches pursued thus far. However, it presents a research gap which can speak to the core components of infrastructure, trade but eventually take into account broader geo-political considerations and formulate international trade as a bottom-up system rooted in the network of trade corridors that allow it to function and can represent interactions in a system-like manner. Such a system could take explore the interaction of micro networked infrastructure observations (looked at by logistics / supply chain work) with mezzo and macro level phenomena (typically examined by the economics researchers). Practically, it is an extension of the calls for further research from the country-level analysis that includes the work of bottom-up approaches which are crystallised below.

2.6 | Literature review synthesis and research questions

What challenges can be seen across the literature? What questions will this work address?

Sections 2.1 through to 2.5, outlined key conceptual and theoretical building blocks, along with focus in different strands of literature and challenges with relation to the BRI that are pertinent to this work. These are outlined in Table 6 with ways they will be address in this work. It is intended for this to provide a series of questions and objectives which underpin the methodology constructed in the next section with the aims of transparency, reproducibility and a figurative 'plug and play' / modular nature for the theoretical as well as the data components gathered.

Table 6 |
Conceptual / research challenges (C) and suggested resolutions

#	Challenge	Highlighted by and discussed in	Suggested resolution
C1	**Absence of systems theory**	e.g. Jervis (2010), Harkavy (2013) and Hodgson (1987) (Sections 2.1 through to 2.5)	**Select methods compatible with systems theory in relevant disciplines.** The fundamental theoretical components, be they in geo-politics, trade or transport, lend themselves to the application of systems. Through the selection of appropriate methods such as network / graph theory a base system of the BRI can be constructed that can be enhanced over time. **Question 1:** How can system theory be used to represent the BRI?
C2	**Open-minded conceptualisation of the BRI**	Yuan (2020) (Section 2.5)	**A framework and analysis which is made with the expectation of change.** This challenge cannot be rectified at present, as the BRI is not conceptualised clearly enough. However, developing a framework and analysis that expects change in assumptions or dataset would make it easier to accommodate evolutions of the BRI over time. **Objective 1:** Outline a clear framework and transparent methodology and code/data approach
C3	**Combination of theoretical lenses**	Yuan (2020) (Section 2.5)	**Incorporation of a number of key dynamics from trade and transport areas.** In combination with the modularity of C2, there is scope to use different theoretical lenses either at different spatial levels of a system / sub-system OR compare / contrast outcomes with different base assumptions. **Objective 2:** Allow for analysis across varying spatial scales
C4	**Non-western perspectives**	Yuan (2020) and Lee et al. (2018)* (Section 2.5)	**Be cognisant and seek out work in an accessible language that may bridge this gap.** In this case, Lee et al. (2018) outline a number of works and categorise them by methodology and focus in order to get a more complete picture of ongoing research agendas. **Objective 3:** be mindful of broader sources of literature
C5	**Incorporation of transport literature**	Baier et al. (2018) (Section 2.3)	**Include theory concepts, as per C3; focus on bottom up analysis that includes key transport nodes.** Mirroring the framework and efforts of authors already presented, generate a bottom up system based around routes and their nodes while being mindful to ensure that the data is accompanied by approaches, concepts and theories that feed into trade and other disciplines in line with C3. **Question 2:** What additional information can be obtained from route-level analysis? **Question 3:** Are results in line with existing literature?
C6	**Lack of data**	Evident throughout literature + Lee et al. (2018)	**Granularity, modularity and clear assumptions for incomplete data.** A bottom up structure is likely to come across data gaps across different countries. In line with Objective 1, a framework and workflow that expects changing datasets and definition of the BRI can contribute in solving this challenge by allowing for assumed values to be utilised until such time as they can be replaced with better estimates or real data. **Question 4: Can a reliable and easily accessible starter dataset be constructed?**

*Note: *Lee et al. (2018) indirectly pointed this out by highlighting their emphasis of non-western literature*

Viewed together, Table 6 above and Figure 3 in the previous section together highlight key questions that arise from the fundamental enquiry revolving around systems theory and BRI specific research agendas respectively.

Overall, the central line of enquiry is how can international trade, or a subsegment like the BRI be represented as a system. What are its inputs, internal mechanics, and outputs? Can a general reference dataset be constructed and maintained so as to provide a common base for future research? Furthermore, what additional information could be learnt, when pursuing research agendas such as modelling cargo flows and identifying new routes? Finally, does this construct corroborate results from existing literature?

It is not possible for one work to address all these questions. However, to serve as a proof of concept and a base for further work, objectives have been defined in response to persistent challenges in the literature. These range from an expectation of missing data and allowing for new findings to be interchanged for assumptions made to allowing any analysis to be conducted across various levels of analysis.

Adding to the above the elusive concept(s) of distance then allows one to investigate specific routing problems as well as the impact of specific infrastructure changes over time.

As a result, the following research questions are specified:

Overarching research questions (RQ):

- **RQ1** | How can the BRI be reconstructed using systems theory?
- **RQ2** | What additional information can it provide?

Specific questions:

- **RQ3** | What is the impact of utilising other measures of distance in a network representation of the BRI compared to the traditionally used geographical distance?
  - a | How may this affect estimates of cargo flows?
  - b | Are any unexpected new routes identified as a result?

The below sections provide an overview of the data landscape and methodologies that will be used address these questions.

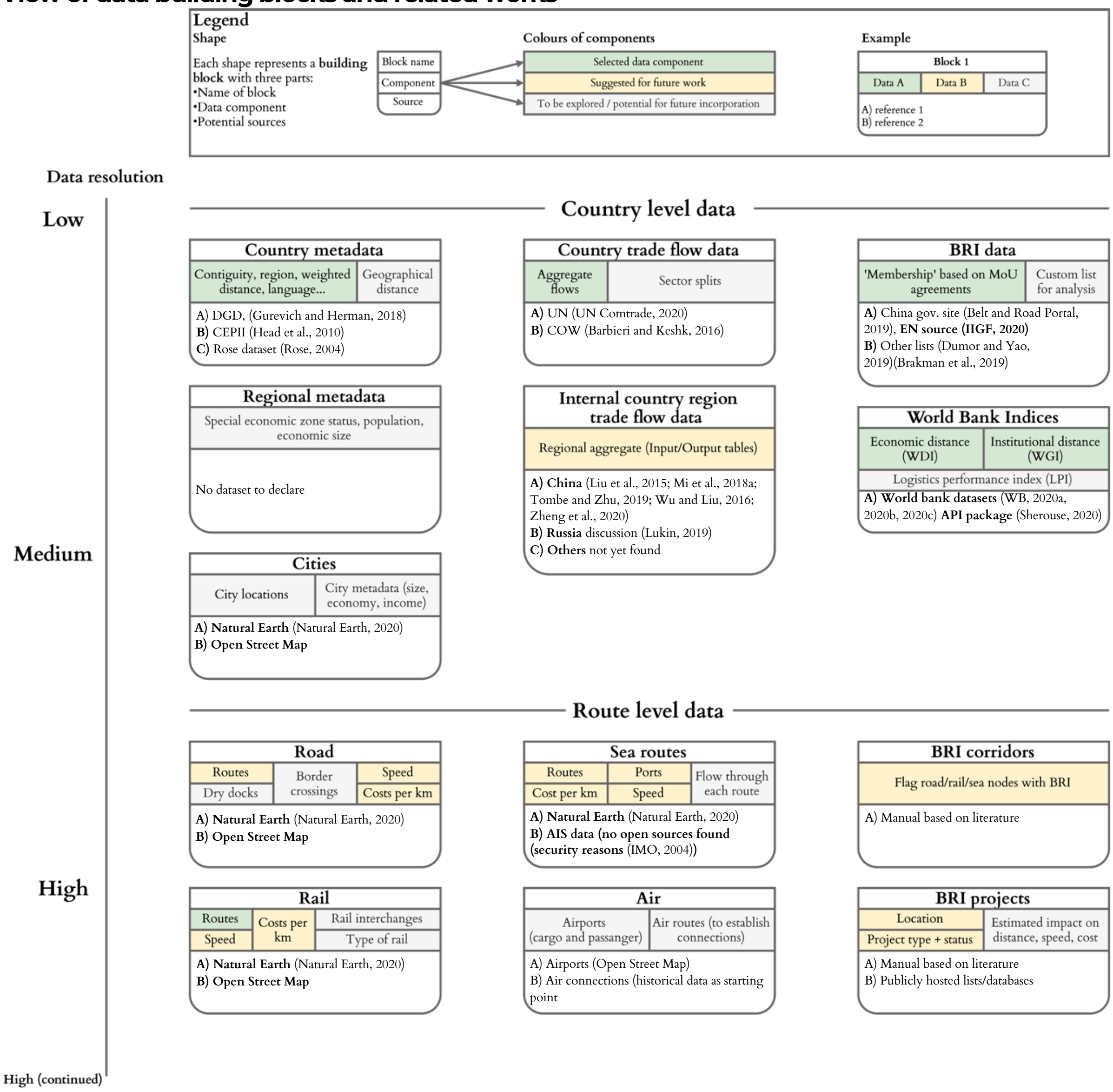
3 | Data

3.1 | Outlining the desired and selected building blocks

What is the full scope of data required? What will this project use for a proof of concept?

Answering the overarching research question (RQ1) is an iterative process with many possible inputs that may evolve or indeed change over time with subsequent research projects. While constructing a proof of concept, it is important to keep in mind that there is a wide array of data 'building blocks' already available and account for the possibility that these may change, be that through updates or replacement by another dataset. As a result, datasets encountered are classified into three categories, those selected for a proof of concept, those not selected but eventually desirable and potential extra or 'nice to have' datasets. Figure 4 outlines a number of data points identified thus far along with potential sources and categorises them in line with the above. Although not an exhaustive list, Figure 4 represents sources typically found in the literature, with further debate below.

Figure 4 |
View of data building blocks and related works



Note: Gravity datasets: (Head et al., 2010), (Rose, 2004), (Gurevich and Herman, 2018); Flow datasets: (UN Comtrade, 2020)(Barbieri and Keshk, 2016); BRI membership: (Belt and Road Portal, 2019) (IIGF, 2020) (Dumor and Yao, 2019)(Brakman et al., 2019) ; Regional data: (Lukin, 2019), (Liu et al., 2015; Mi et al., 2018; Tombe and Zhu, 2019; Wu and Liu, 2016; Zheng et al., 2020); Cities: (Natural Earth, 2020a); Sea routes: AIS data typically not openly available due to security reasons (IMO, 2004), some starting points include (OSM, 2020). Air data: (OpenFlights, 2020); World Bank, access via API (Sherouse, 2020), datasets: (WB, 2020a, 2020b, 2020c)

3.2 | Low-resolution datasets

What was considered in selecting and processing these datasets?

Datasets categorised as 'low' resolution represent datasets that aggregate information to a country level. Commonly found in international trade literature, there are so-called 'gravity datasets' that capture country metadata found to have explanatory power in understanding trade. These variables include elements such as language, international organisation membership, pure geographical distance or in some case weighted distance measures (weighted by population density or city location)(Mayer and Zignago, 2006). Three datasets have been observed to be used in the literature, the Rose dataset by Rose (2004), CEPII dataset by Head et al. (2010) and the Dynamic Gravity Dataset (DGD) by Gurevich and Herman (2018). For this work, the Gurevich and Herman 'DGD' dataset is chosen. Firstly, and most importantly, it is the most up to date spanning 1948-2016. In contrast, Rose and CEPII go up to 2004 and 2014/15 respectively. Secondly, the 'dynamic' part of the dataset allows for more robust tracking of territories and their colonial histories. Although this is not a focus of this work, it is useful information to have when studying longer-term trends.

Table 7 |
Dynamic Gravity Dataset (DGD) overview

Type of field	Fields	Used for
General	Year	Kept to 2016 for proof of concept, can be expanded for broader range by adjusting filtering
Destination country data	'country_d', 'iso3_d', 'dynamic_code_d', 'landlocked_d', 'island_d', 'region_d', 'lat_d', 'lng_d'	'iso3_d' serves as key for matches, descriptors such as 'landlocked_d', 'island_d', 'region_' used for filtering
Measurement	'gdp_pwt_const_d', 'pop_d'	GDP in constant dollars and population count, retained for reference
Origin country data	'country_o', 'iso3_o', 'dynamic_code_o', 'landlocked_o', 'island_o','region_o', 'lat_o', 'lng_o'	Mirror of 'destination' data but for countries where connection originates
Measurement	'pop_o'	
Paid information	'contiguity', 'distance',	Flag to determine whether two countries share a border and therefore would be connected and the distance measure between them

Note: This table reflects fields used, for full description of dataset see Gurevich and Herman (2018)

Complementing the gravity datasets are 'trade flow' datasets that measure the bilateral trade flows at an aggregate or sector level between countries. Ideally, sectoral breakdowns would be included, they are omitted for the moment with the focus on aggregate import/export flows. Prominent sources of this data include the United Nations Comtrade dataset (2020) and the Correlates of War 'COW' outlined by Barbieri and Keshk (2016). As with the gravity datasets, the coverage of UN Comtrade to 2019 combined with regular updates and the potential for studying monthly data resulted in it being chosen. Furthermore, it is easily accessible via public API.

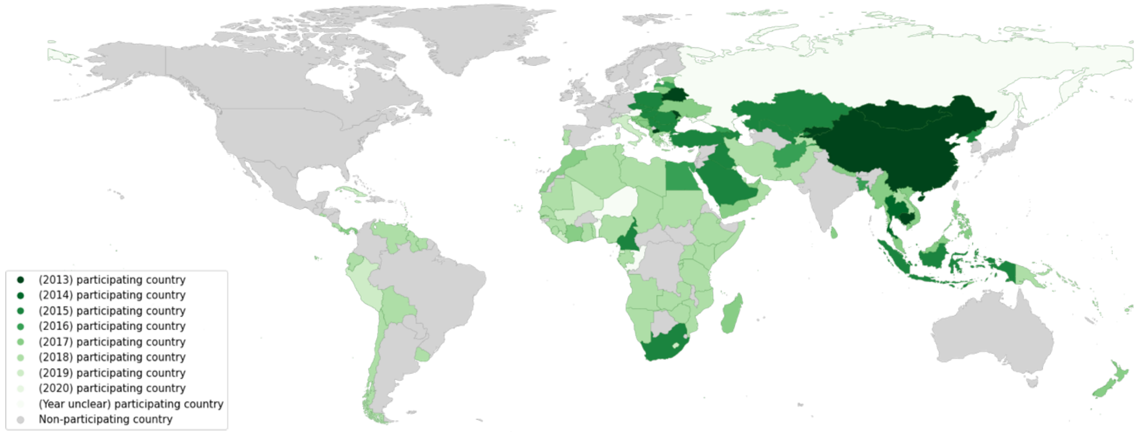
Table 8 |
UN COMTRADE dataset overview

Field	Description	Used
rtCode (int)	UN dataset code for this reporting country (country submitting data)	Retained for reference to source
rt3ISO (string)	ISO3 code for country, used for matching with other datasets	Used for joining other datasets
ptCode (int)	UN dataset code for a partner country	Retained for reference to source
pt3ISO (string)	ISO3 code for country, used for matching with other datasets	Used for joining other datasets
rgDesc (string)	Classification of 'export', 'import' 're-export' classifiers	Exports kept
yr (int)	Year column, used for matching alongside ISO3 codes	Used for joining other datasets
Value (int)	Dollar value of measure in 'rgDesc'	In this case, we focused on exports

Note: This table reflects fields used, for full description of dataset see UN Comtrade material.

Finally, a variable needs to be created for the purpose of denoting the membership of the BRI. As discussed in Section 2.4, this is ambiguous at best and varies from work to work, with some authors focusing on routes as a more concrete indicator. Nevertheless, an entry ought to be made. A possible, yet admittedly imperfect method is through taking the signing of Memoranda of Understanding (MoU) and similar as proxies for membership / cooperation available on Chinese government portals (Belt and Road Portal, 2019) and other publicly facing portals (IIGF, 2020). Unfortunately, as there are sometimes, multiple MoUs signed or not always consistently reported, research as part of this work ended up making some corrections to the IIGF list . This did not change the overall count of member countries (138) but it affected the year they are understood to have joined.[8] Visualised, membership can be seen to have expanded from the immediate neighbourhood of China and steadily moved outwards.

Figure 5 |
BRI participation over time-based on signed Memoranda of Understanding



The low-resolution datasets are relevant and important to correctly assign attributes to higher resolution datasets and to be able to consistently aggregate analysis for comparison to other works and formulation of various interactions into systems.

[8] Amended dataset is available alongside this work at the GitHub repository here: https://github.com/antoniosfiala/UCL_Dissertation

3.3 | Medium-resolution datasets

What was considered in selecting and processing these datasets?

Moving from low to medium resolution datasets, there are two datasets that are desirable but omitted at this stage, regional/province metadata and regional/province trade flows. The value add of these datasets speaks to a more refined understanding of flow origin points. For instance, instead of taking Country X with 1,000 exports, taking 5 provinces with distributed exports can speak to the specific distance they face in order to reach an end destination. The principal challenge is availability, particularly around the internal country trade flows. For example, a number of works referenced utilise Chinese regional input/output tables to assess flow of trade and development over time, however the base data is available, intermittently with recent cases being 2012 and 2015 (Zheng et al., 2020). In other instances, i.e. Russia these are not produced (Sayapova, 2020). Intuitively, these would be crucial additions to a refined view of trade dynamic across the BRI, however for now they will be omitted with a means for their inclusion in the future.

The above provides data that can be used in its own right for country-level analysis. However, in this case, it can be used as metadata that gets assigned to more detailed entries that can form part of a network of routes. Cities are one such dataset which is available in the public domain and can serve as a mezzo layer of connection. Open data sources include Natural Earth or Open Street Map (OSM) which contain data on names and locations (Natural Earth, 2020b; Open Street Map, 2020). Although the cities can inherit metadata from the national or regional level described above, their own metadata such as population and size of economic remains challenging. For population, some entries are available but as per OSM's own site, only ~ 36% of entries have a population marked, other information will be harder to obtain or will have to be acquired on case by case basis. Secondly, having the cities does not directly tell us about whether they are linked and the nature of this connection, such as its quality, distance and other factors that could result in 'trade friction' (tolls, border crossings).

Although neither dataset will be used for this proof of concept, accounting for their existing and eventual incorporation is important in order to adhere to the objectives stated for this work.

3.4 | High-resolution datasets

What was considered in selecting and processing these datasets?

Datasets categorised as 'high resolution' help solve the 'connection' issue between points such as cities while increasing the number of nodes that could be of interest. From a general transportation perspective, there are modes over land, air and water/sea (larger rivers/lakes can be used too). Land is typically segregated into road and rail, while air and water are largely self-contained categories. The BRI affects all, with particular attention to land and sea, however as a starting point the focus here will be on rail connections as they carry the majority of overland trade. In keeping with the principles of this project, other types of connection and points along the network such as ports and airports are expected additions down the line and so the data extraction and preparation code accommodate this.

It is important to concede that by omitting sea routes, a very selective view of connectivity is presented. This limitation is the core reason why this is positioned as a proof of concept that explores the utility of this approach as opposed to one that will lead to the actionable conclusions at this stage.

In addition, there are specific BRI high resolution data points. These include, tagging individual projects, groups of nodes as belonging to scope of a project or an overall transport corridor (spanning vast distances) that is the focus of a particular study. Datasets such as this appear to exist given the existence of work from a number of authors (see Table 5), however datasets accompanying academic research focused on projects are not typically publicly available. There are some publicly available lists, as well as research groups such as Mercator Institute of Chinese Studies (MERICS) who maintain their own version in-house. Consequently, it is acknowledged while some data exists out there, it is omitted for this study and encouraged to be an evolutionary step.[9]

These datasets are crucial in allowing the bottom-up analysis at the heart of this endeavour that can be contrasted to outputs of lower resolutions datasets and serve as the source based on which connectivity is established and develops in the system.

[9] See appendix for details on interaction with MERICS and suggested open lists.

3.5 | Extraction, processing and storage

How are selected datasets extracted, processed and stored?

Extraction of the data is carried out using Application Programming Interfaces (API) calls where possible or extraction of raw data in excel/.csv before processing in Python. Raw data is kept as is while processed data (typically as Panda data frames) are stored as compressed comma separate value files and stored on GitHub. Table 9 goes into more detail for each data input.

Table 9 |
Dataset processing summary

Dataset	Denoted as	Access via	Stored as
DGD	$Data_{grav}$	.csv file access directly from host site	Compressed gzip csv file
UN COMTRADE	$Data_{UN}$	API (call created by author in python)	
World Bank	$Data_{WB}$	API (using python package)	
BRI membership	$Data_{BRI}$	Compiled by author, accessed from local .csv file	.csv file
Rail routes	$Data_{NE}$	Downloaded shape file from Natural Earth site	.shp file

At this stage, the author notes that the DGD data had been amended to correct some omissions (such as shared Greek-Bulgarian border) and change other values around contiguity so that they better reflect the geographical reality of where connections between countries are likely to be used as corridors.

The high-resolution dataset for rail routes has not been subject to changes / corrections at this stage.

3.6 | Ethical considerations

Are there ethical considerations based on the datasets used or findings generated?

The datasets in this research are all sourced from public sources, which are clearly cited. The findings speak to high level observations on connectivity between countries as a whole and the rail routes between. No ethical risk is expected.

4 | Methodology

4.1 | Method selection and key concepts

What methods are selected for analysis and what assumptions/decisions are made?

Addressing RQ3 and building the network is a problem that in line with the BRI literature and broader literature lends itself to the use of graph theory to construct a network that represents the connections within the BRI. The terms 'graph' and 'network' are used interchangeably depending on the works referenced at that point of writing. It is acknowledged that while nuanced differences exist, these are not consistent enough to strictly adhere to only one term. This section will touch on the use of networks in the broader literature as well as take the time to explain key elements of the networks and how they will be used.

Graph theory is a versatile toolkit deployed in applications that represent relationships, connections or dependencies in processes. Nystuen and Dacey (1961) provide a helpful guide for what phenomena are suitable by stating:

> *"A particular phenomenon is suitable for this type of analysis when it may be viewed as a relationship or flow that links objects that are properly mapped as points." (Nystuen and Dacey, 1961, p. 30)*

In the context of the BRI, the additional benefits of applying graph theory to construct a network are in asking questions around resilience and identification of weak points as the work of Wagner and Neshat (2010) or Masad (2014). Furthermore, although not observed in the study of the BRI, there are works the examine the deployment of networks in terms of international trade and complex systems (see Vidmer et al. (2015) or Li et al. (2019)) as well as, construct multi-level networks (see Oselio et al. (2014)). See Table 10 for more detail.

Table 10 |
Sample of relevant graph theory application

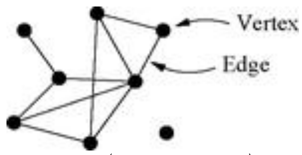
	Citation	Graph theory	Trade	Resilience	Complexity	Comment Topic / Problem addressed by the paper | Key finding of research | Relevance of the work to this project
1	(Nystuen and Dacey, 1961)	✓	✗	✗	✗	**Topic / Problem** | Network hierarchy construction based on flow (focused on cities) **Finding** | Sharing a technique to achieve this goal, recommendation to deploy on states **Relevance** | Equivalent creation of such hierarchy for BRI can be used to describe the created network
2	(Oselio et al., 2014)	✓	✗	✗	✗	**Topic / Problem** | Analysis multi-layered networks **Finding** | Introduction of algorithm for finding approximate pareto front **Relevance** | Later iterations are likely to satisfy multilayer condition where this algorithm could be used
3	(Vidmer et al., 2015)	✓	✓	✗	✓	**Topic / Problem** | Complex system evolution prediction **Finding** | New metric developed on which new link prediction shown to be more effective **Relevance** | View of international trade as complex network system
4	(Li et al., 2019)	✓	✓	✗	✓	**Topic / Problem** | Use complex network analysis to show evolution of robotic trading over time **Finding** | Network density increasing suggesting diversification **Relevance** | Comparable study by way of scope, good reference for types of analysis to carry out
5	(Masad, 2014)	✓	✗	✓	✗	**Topic / Problem** | Resilience within a network in reaction to stochastic shocks **Finding** | Identified specific countries which are most likely to be affected by oil shocks **Relevance** | Use of networks to study resilience, transferable to transport infrastructure
6	(Wagner and Neshat, 2010)	✓	✗	✓	✗	**Topic / Problem** | Tackles application of graph theory in resilience which is atypical **Finding** | Approach for measuring and managing supply chain vulnerabilities **Relevance** | Showcases limitation of complex systems based on high data requirements

From Nystuen and Dacey, we showed *where* a network can be suitably deployed. Newman (2018), through his textbook, speaks to the nature of *what* a network is. In his words:

> *"A network is a simplified representation that reduces a system to an abstract structure capturing only the basics of connection patterns and little else."*

There are two basic components to consider, nodes and edges. A node (sometimes called a vertex) represents an entity or point of interest such as a country, city, port or border crossing. An edge represents a relationship between two points. This can be a physical connection over land, a route connecting two ports (be they air or sea) or a softer connection denoting commonalities such as share language, diplomatic ties and social connection. In this work, there are two kinds of connections, an abstract connection between countries and a physical connection determined by the rail routes that run between countries.

Figure 6 |
Basic network diagram from Newman (2018)



Source: (Newman, 2018, p. 10)

The creation of systems based on these two building blocks is a balance between the reduction and retention of complexity, as well as the scale / coverage of the network. In this case, we are aiming to build a proof of concept from the bottom up based on multiple sources. Therefore, while maintaining some complexity, other elements are left to further research. The key to creating a more robust representation of the BRI is the utilisation of the data blocks discussed above and highlighted in Figure 4 to define attributes of the network's nodes and edges presented in Table 11. These include labels, but also strengths – sometimes called 'weights' – that allow one to define and store more information about a system (Newman, 2018). In this case, the weight attribute is what our distance variable at the core of RQ3 can be captured by. Analysis can then be run on the differently weighted networks.

Table 11 |
Defining attributes of the BRI networks for visualisation and analysis

Component	Attributes (vis. / analysis)	Country *(dataset)*	Route *(dataset)*
Node	**Colour** (visualisation)	Red for highlights, else follow measure colour	Highlight based on measure colour
	Position (visualisation)	Country lat/long centroid *($Data_{grav}$)*	Lat/Long of given node along route *($Data_{NE}$)*
	Size (visualisation)	Constant except for layered nodes	
	Shape (visualisation)	Constant (circle), future extension can designate BRI membership *($Data_{BRI}$)*	
	Labels (visualisation)	Country name *($Data_{grav}$)*	None
	Additional parameters	None	All relevant country level information, such as country name, LPI measure etc.
Edge	**Connection** (analysis)	Geographical contiguity *($Data_{grav}$)*	From raw data *($Data_{NE}$)*
	Colour (visualisation)	Use to highlight shortest route	
	Weight (analysis)	Distance *($Data_{grav}$)*	Distance *($Data_{NE}$)*
	Size (visualisation)	Constant except for emphasis of routes, future extension can show extent of trade flow *($Data_{UN}$)*	
	Shape (visualisation)	Solid	
	Labels (visualisation)	None	None
Direction		Unassigned. No direction, all trade connections are assumed to have the capacity to be bidirectional. Future iterations could deploy bidirectional import/export flows	

Note: Table arrangement based on work of Ognyanova (2019)

With the prominence and history of networks, there is an accompanying toolkit, typology and concepts that help analyse networks. Table 12 outlines key concepts and reflects on their relevance for this work.

Table 12 |
Network concepts based on Newman (2018) and Nystuen and Dacey (1961)

Concept	Explanation	Relevance
Degree	Number of edges/connections a node has	Use to gain insight into global or local characteristics of a network and provide a starting point for describing it.
Path	Sequence of nodes that go one point to another. There are different types such as geodesic paths (shortest)	Can assess changes in identifying the geodesic path from origin to destination when the measure of distance changes in a network.
Component	A subset of a network containing a series of nodes and edges	No direct relevance right now, but may be useful in identifying sub-systems which may cause the addition of unnecessary distance or friction to a path
Transitivity	There is a transitive property where in a hierarchy positions are passed on. An example from Nystuen and Dacey (1961) outlines that if *a* is subordinate to *b* and *b* is subordinate to *c*, it logically follows that a is also subordinate to *c*.	Subordinate status can be identified in a number of ways, Nystuen and Dacey focused on flows of people, in this case, the flow of goods can be used.
Centrality	A measure that helps identify the most or a set of most important nodes in a network. Measurement methods vary from using degrees to using flows. Typically, it is on this basis that hubs are identified in a network.	As a fundamental concept of networks, it relevant in describing any network or subnetwork constructed. Although not expected to change with changes in distance (as the number of edges remains the same) when based on degree counts, it can be affected by the combination of different distances and flows should those affect the paths taken.
Betweenness centrality	Betweenness centrality "measures the extent to which a (node) lies on paths between other (nodes)." (Newman, 2018, p. 187)	Another critical concept used to provide an overview of networks and can help identify potential influential nodes or those critical from a resilience point of view. In the context of international trade, one can consider the Suez or Panama canals to have a high 'betweenness centrality' value due to their volume of traffic from different routes that pass through them.

Modelling geographically constrained elements such as rail infrastructure as a network, expectations can be developed based on certain traits of the network itself. For instance, as illustrated by Cardillo et al. (2006) in their work examining street networks, their finding that the degree distribution of 1-5 connections per node can be taken as an expectation through which to test any networks created. Provided the rail network exhibits comparable properties to the streets, further analysis could examine transferring city centred approaches to another type of network.

Finally, in order to test the concepts described in Table 12, a number of distance variables need to be used in order to assess their impact on concepts such as identifying the shortest path, and betweenness centrality.

4.2 | Working with distance

What is being measured and how?

With the data in place and a structure of the network outlined, this section focuses on defining the types of distance that will be used iteratively to create differently weighted BRI networks (highlighted in Table 11). Section 2.3 along with Table 2, touched upon the existing debates with regards to 'distance' in international trade. This sections builds on the contributions to this debate by Wu et al. (2020) and Wang et al. (2020) who both leverage works across a number of fields and incorporate four types of distance into their investigations. These are, geographical/spatial (SD), economic (ED), cultural (CD) and institutional (ID) distances, with Wu et al. (2020) going a step further to propose a composite measure made up of all four referred to as National Distance (ND).

Wang et al. (2020), systematically describe their process, formulas and data inputs in the appendix accompanying their paper and is recommended for an in-depth overview of their methodology and supporting research. For brevity, this work only summarises relevant parts and formulae of a selected subset of measures in Table 13 where *i* and *j* represent pairs of countries or nodes along rail routes. Any adaptations made to the original approach are discussed below as each distance type is discussed in turn.

Table 13 |
Defining distance measures

Distance type *(dataset)*	Distance indicator metric	Symbol	Approach / formula
Baseline *(default value)*	Take each edge as having value of 1	D^B	$D^B_{ij} = 1$
Simple geographical / spatial distance *(Data$_{NE}$* (Natural Earth, 2020b))	Applies to route level analysis, actual geographical distance covered by a segment of a route	D^G	$D^G_{ij} = distance_{ij}$ Where i and j represent a pair of nodes forming an edge and distance is the given value in the *Data$_{NE}$*
Weighted geographical /spatial distance (WD) *(Data$_{grav}$ (Gurevich and Herman, 2018))*	Weighted greater circle distance, pre-calculated in dataset.	D^{GW}	Greater circle distance weighted by population and location of large cities. (Gurevich et al., 2018, p. 13) – more detail also in Appendix
Logistics adjusted distance *(Data$_{NE}$ + Data$_{WB}$)*	Applies to route-level analysis	D^{GL}	Original LPI index (ranging from 1-5, 5 is the best) is rescaled to a range between 0 and 4 denoted as *LPI_{base}*. After that the index is translated into a metric indicating a penalty (*LPI_p*) as per this formula: $LPI_{pij} = \frac{LPI_{base}}{4} + 1$ Penalty is applied for routes: $D^{GL}_{ij} = \mathrm{LPI}_{pij} \times distance_{ij}$ An average between countries taken for countries: $D^{GL}_{ij} = \frac{\mathrm{LPI}_{pi} + \mathrm{LPI}_{pj}}{2} \times distance_{ij}$
Economic (ED) *(Data$_{WB}$ (WB, 2020b))*	Difference in per capita real GDP ($\bar{y}$)	D^E	$D^E_{ij} = \lvert \bar{y}_i - \bar{y}_j \rvert$
Institutional (ID) *(Data$_{WB}$ (WB, 2020a))*	I_1 : Control of corruption I_2 : Government effectiveness I_3 : Political Stability and Absence of Violence I_4 : Rule of Law I_5 : Regulatory Quality I_6 : Voice and Accountability	D^I	$D^I_{ij} = \sqrt{\sum_{p=1}^{n} [\frac{(I_{pi} - I_{pj})^2}{V_p}]}$ p is the number of each I dimension Using Euclidian distance formula instead of KSI index formula (Konara and Mohr, 2019)

Note: This table recreates in large parts that of Wang et al. (2020) – Table A2-1 in their appendix; i and j represent pairs of countries ; $\bar{y}$ represents GDP ; p is the number of each I dimension (1-6)

There are two points of departure from the works of Wang et al. (2020) and Wu et al. (2020). Firstly, the complement of measures is different. The focus is on datasets that are accessible and periodically and reliable refreshed. This makes cultural (CD) distance less desirable and leaves economic (ED) and institutional (ID) distances as measures that are retained. However, in contrast to these authors, more emphasis is placed on geographical distances and other factors that affect the routes.

To begin, a baseline for any network constructed will be assuming all edges to be of equal length and is denoted by D^B. This is expected to generate results that identify paths with the fewest edges along the way and serve as a frame of reference to consider against when the remaining measures below.

Next, there are three measures that are varieties of geographical distance. Firstly, a simple, point to point geographical distance (D^G) derived from rail route data in $Data_{NE}$ with the use of QGIS software. D^G is only utilised for route-level analysis. Secondly, a weighted measure of distance utilised for country distances (D^{GW}) taken from $Data_{grav}$. This is the typical measure utilised in gravity modelling and consists of a country centroid coordinate that was derived through a calculated great circle distance approach (see appendix for more detail). Thirdly, a composite distance measure which takes into account the quality of logistics (D^{GL}) based on Logistics Performance Index (LPI) available as part of $Data_{WB}$. This measure is utilised across country and route spatial scales and can include a temporal dimension. Time is not considered in this work but can be a further avenue of exploration.

The second point of deviation from the work of Wang et al. (2020) and Wu at al. (2020) is the method of calculating distance for the remaining measurements. This project utilises Euclidian distance in contrast to the original formula typically used to measure distance along the cultural and institutional dimensions. The rationale is driven by the work of Konara and Mohr (2019) who demonstrate that the original formula leads to exaggerated measurements at greater distances due to the original taking an average of the square units instead of utilising non-square units. The different formulations can be seen below.

$$D_{ij}^{I} = \frac{\sum_{p=1}^{n}[\frac{(I_{pi} - I_{pj})^2}{V_p}]}{n}$$

(original formula)

$$D_{ij}^{I} = \sqrt{\sum_{p=1}^{n}[\frac{(I_{pi} - I_{pj})^2}{V_p}]}$$

(revised based on Euclidian distance formula)

D^E goes beyond the simple difference in income per capital and utilises a pure Euclidian distance measure (Barrett, 2005). The rationale being that the deviation matters rather than the directionality (i.e. positive or negative). We can expect this to lead to situations where paths among similar countries are made ‘shorter’. This point will be re-examined for its appropriateness in the context of this study in the analysis section.

Finally, the D^I is a measure where further to the formula divergence, Wu et al. (2020) and Wang et al. (2020) differ between themselves. The difference for them lies in the dataset of choice. While Wang utilises data from the Heritage Foundation, Wu leverages World Bank data. Given its accessibility the choice was made to follow Wu and prioritise accessibility.

Each group of measures serves a different purpose when it comes to comparisons. For both country and route analyses, D^B establishes a baseline against which other findings of the same scale can be compared. For the construction of country networks, D^B, D^{GW}, D^{GL}, D^E and D^I are calculated and compared. For routes, D^B, D^G and D^{GL} are the measures of choice. In each instance, every measure describes certain dynamics of each network with the geographic distance-based distance allowing comparison across scales.

4.3 | Experimental framework

How will the research questions be answered?

Section 4 begins with by stating the aim of outlining the chosen methodology through which to investigate the impact of distance as per RQ3. However, the implications are broader as this process also describes the data assets which are believed to be of use in the construction of a bigger system. In order to address RQ2, it is not only the question of distance by the level of detail that is being considered and what additional insights can be drawn from the utilisation of 'high-resolution' datasets such as rail routes. Together these outcomes then allow for a discussion to take place in reflecting on the utility and the way through which the BRI can be reconstructed as a system.

To answer the direct question of impact of different distance measures on a network representation of the BRI the following steps are proposed as an experimental framework.

1 | Select scope: in this case Europe and Asia (focusing on the 'Belt' part of BRI)

2 | Country-level network
  a. Construct country-level network
  b. Assign distances to edges (D^B, D^{GW}, D^{GL}, D^E and D^I)
  c. Carry out tests along the network as per Table 14, in particular testing routes from China to Germany

3 | Route-level network
  a. Construct route-level network
  b. Assign distances to edges D^B, D^G and D^{GL}
  c. Carry out tests along the network as per Table 14, in particular testing routes from China to Germany

Using a combination of the network fundamentals based on Newman (2018), Cardillo et al. (2006) and the example of Li et al. (2019) networks 2 and 3 will be tested for the measures outlined in Table 14.

**Table 14 |
Assessing networks**

Concept	Calculation	Description	Purpose
Degree	Count number of edges each node has	Expect the number of edges/connections a node has to be within a relatively stable range (in line with Cardillo (2006))	Use to gain insight into global or local characteristics of a network and provide a starting point for describing it.
Shortest path	Addition of distances between nodes	Sequence of nodes that go one point to another. There are different types such as geodesic paths (shortest)	Can assess changes in identifying the geodesic path from origin to destination when the measure of distance changes in a network.
Betweenness centrality	$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t\|v)}{\sigma(s,t)}$ (sourced directly from (NetworkX, 2020)	Calculate proportion of the time a node is along a shortest path between two nodes.	Study the change with different distances.

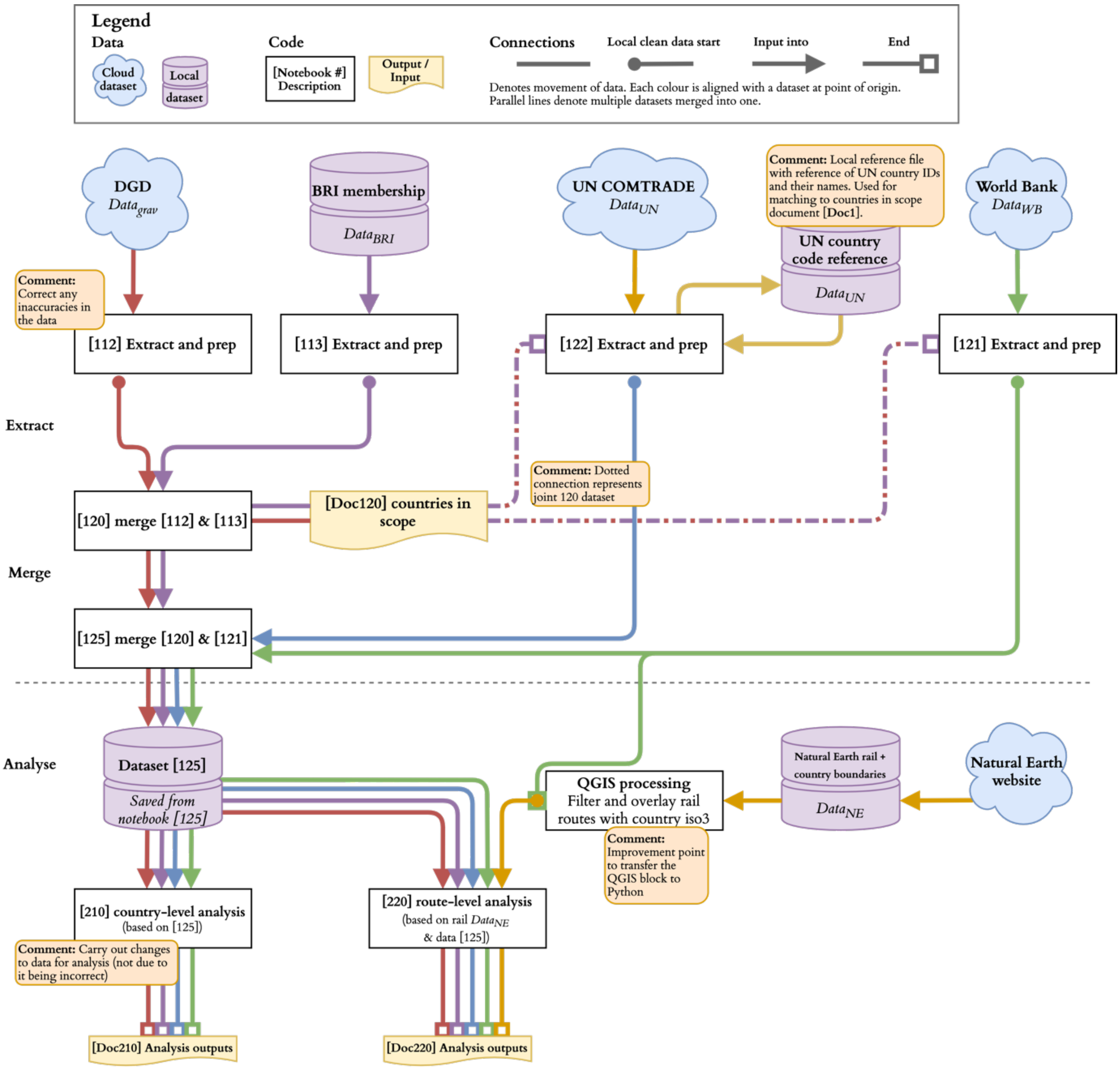
Notes: NetworkX python package algorithm for betweenness centrality based on Brandes (2001)

4.4 | Process overview

How do all the pieces of code and data fit together?

In closing this section, Figure 7 visualise the data flow process and the relevant code stored in Jupyter Notebooks in a GitHub repository. The code and datasets are published alongside this work and are an integral part of reproducibility and expandability. This is an important section in satisfying the stated Objective 1 in Table 6 (Outline a clear framework and transparent methodology and code/data approach). Furthermore, it also acts as a guide / map as to where additions of data or improvement in the code can be made.

Figure 7 |
Data processing flow chart



From Figure 7, there are several points that ought to be expanded on. Firstly, the current data inputs are a variety of local and cloud-based datasets. All country related datasets are ultimately merged in Notebook 125 and stored in a corresponding dataset locally. This is a point where additional information is best added as it allows for a clean cascade down into further analysis, be that country-level analysis in Notebook 210 or additional information for route-level analysis in 220.

5 | Analysis

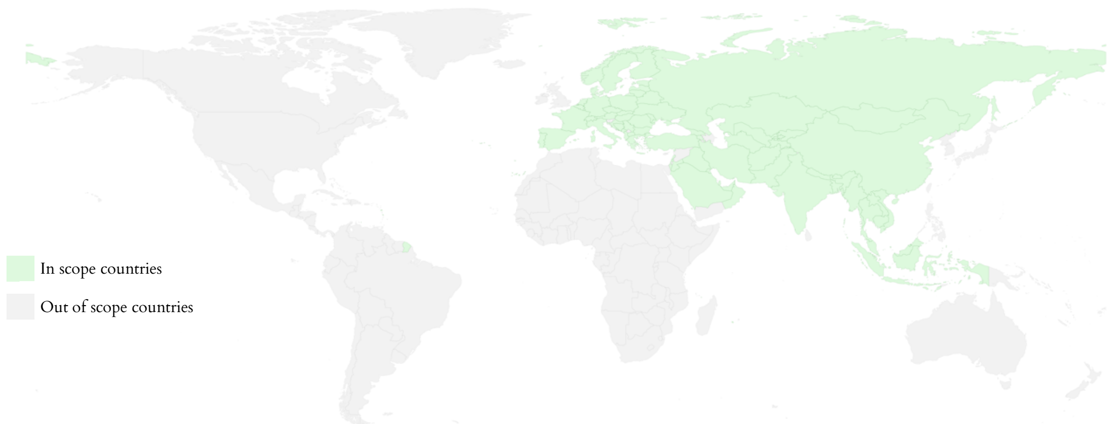
5.1 | Overview

What is the geographical scope? What does the source data look like? How are results presented?

This section carries out the method outlined above in two parts. Firstly, a country-level network construction and analysis of three tests with different distance measures used. Secondly, a route-level (rail) network construction and analysis of the three tests with the distance measures allocated to this analysis.

As per Section 4.3, the geographical scope for this proof of concept are the continents of Europe and Asia. Extracting and filtering the country-level datasets first results in coverage of the majority but not all of the continents. Countries left out were removed either because of inconsistently available data (i.e. Yemen and Syria) or the fact that they were islands and in the interest of simplicity the focus is purely on land-based connections.

Figure 8 |
Selected countries (world map)



Note: There are 70 countries in scope. Countries such as Azerbaijan, Yemen, Syria, North and South Korea, San Marino, Lichtenstein…[complete] were dropped due to lack of data or lack of connectivity once countries who were a bridge link were identified. Further, any island nations from across Europe and / or Asia were dropped.

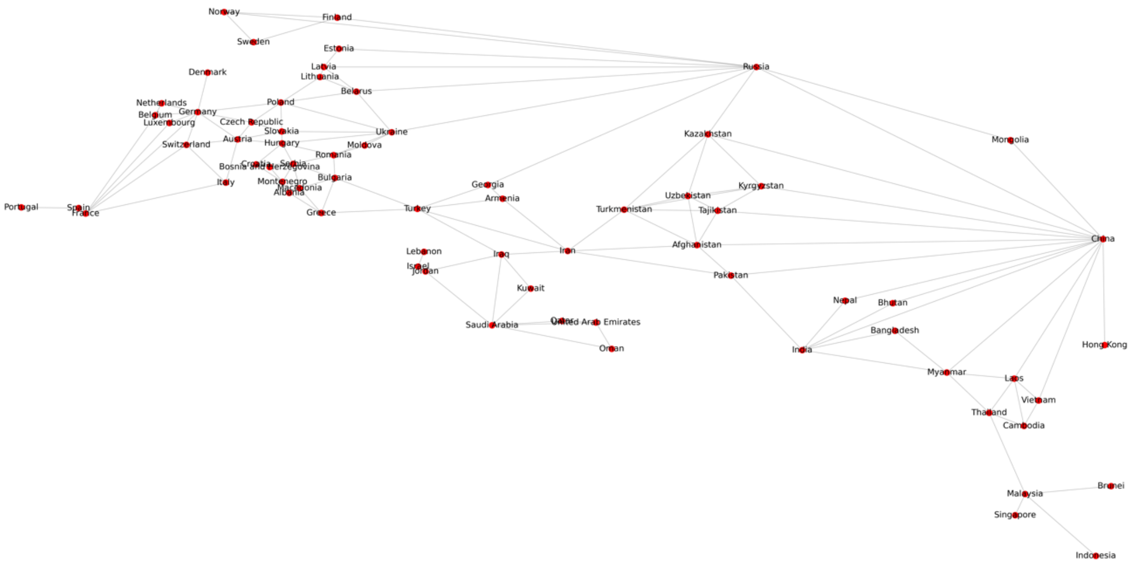
The base layer of data can be found and assessed as part of Notebook 125. Comparative analysis between country and route level analyses is kept to a minimum and retained for the discussion portion of the paper.

5.2 | Country-level networks

What are the findings for country-level networks?

The results presented in this section are all from Notebook 210, which takes the merged dataset [Doc125] and contains datapoints from $Data_{grav}$, $Data_{UN}$, $Data_{WB}$ and $Data_{BRI}$. Constructing and visualisation the network, using the coordinates from $Data_{grav}$ generates Figure 9.

Figure 9 |
Visualising the entire country-level network



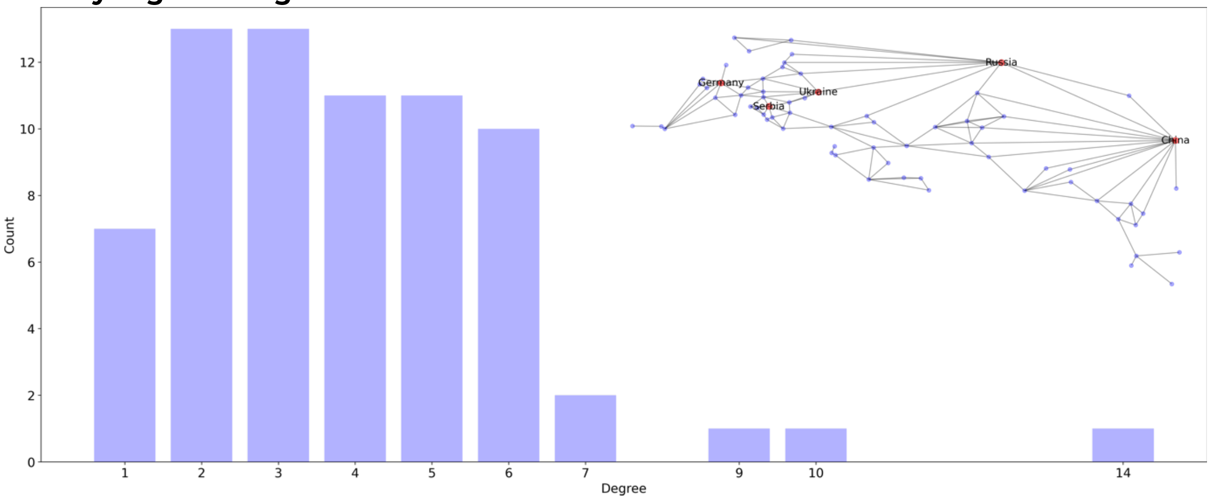
Note: Network created on the basis of border contiguity. Contains 75 countries and 151 edges

On this basis, the remaining test are conducted, starting with understanding the distribution of degrees for each node in the network through Test 1.

5.2.1 | Measuring network degrees

To understand the network of countries, we begin with assessing the degree of each node in the network. In other words, the number of connections each node has. This helps us understand potential key nodes in our network and the distribution, as demonstrated in Figure 10.

Figure 10 |
Country degree histogram and network reference visualisation



Nodes with high degrees are Serbia (7), Ukraine (7), Germany (9), Russia (10), and China (14). This result is largely in line with expectations. Russia is demonstrably an important transition country of the Eurasian land bridge corridor and Germany is its typical endpoint due to its economic size and reliable connectivity to the rest of Europe.

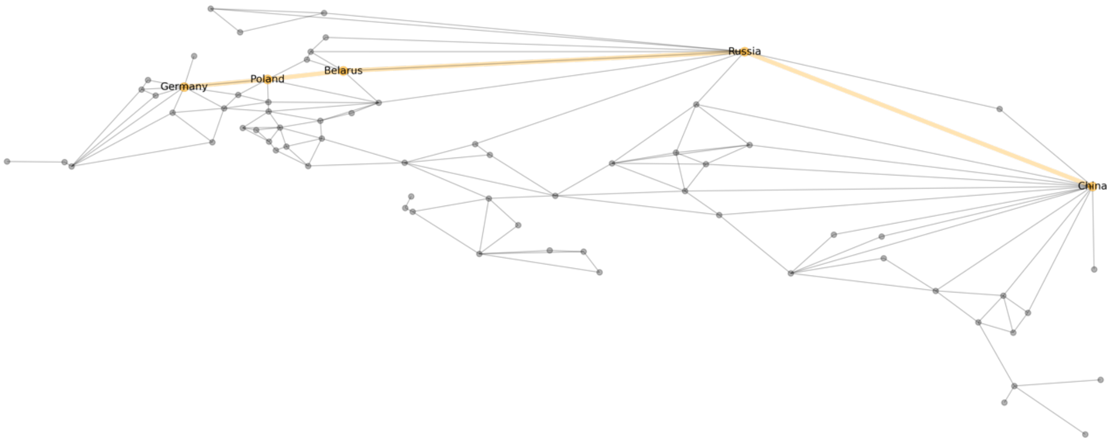
Considering the distribution of results, the majority are in line with the observations by Cardillo et al. (2006) whose work suggests a range of 1-5 for street level networks. In this case, given the number of outliers and the different geographical scale, it is understandable why these types of networks are unlikely to be directly comparable or be able to use equivalent assumptions.

Nevertheless, an expectation to test in the following section is whether these countries find commonly find themselves along the shortest path between common origin and destination nodes using a simple 'shortest path' test and later a 'betweenness centrality' test.

5.2.2 | Shortest path analysis

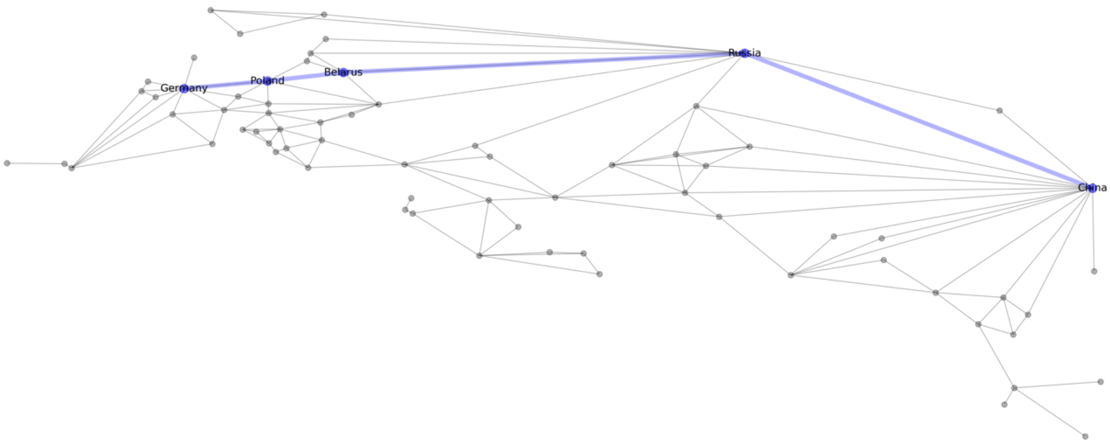
With a constructed network, a basic understanding of its shape and degree distribution this analysis applies the distance measures introduced in Table 13 (D^B, D^{GW}, D^{GL}, D^E and D^I). China and Germany are chosen as starting and end nodes respectively. China as the largest exported in the BRI and Germany as the significant economic actor in the European Union and a large trading partner for both imports and exports. In reality, this route runs from China, through Kazakhstan, across Russia into Belarus before finally crossing Poland in Germany.

Figure 11 |
Country shortest path (D^B)



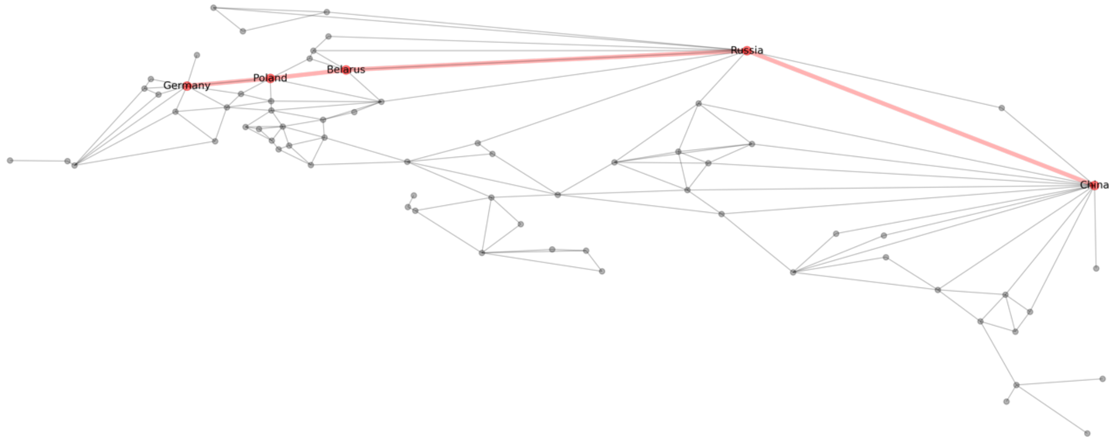
Number of edges: 4
Distance covered: 4 edges

Figure 12 |
Country shortest path (D^{GW})



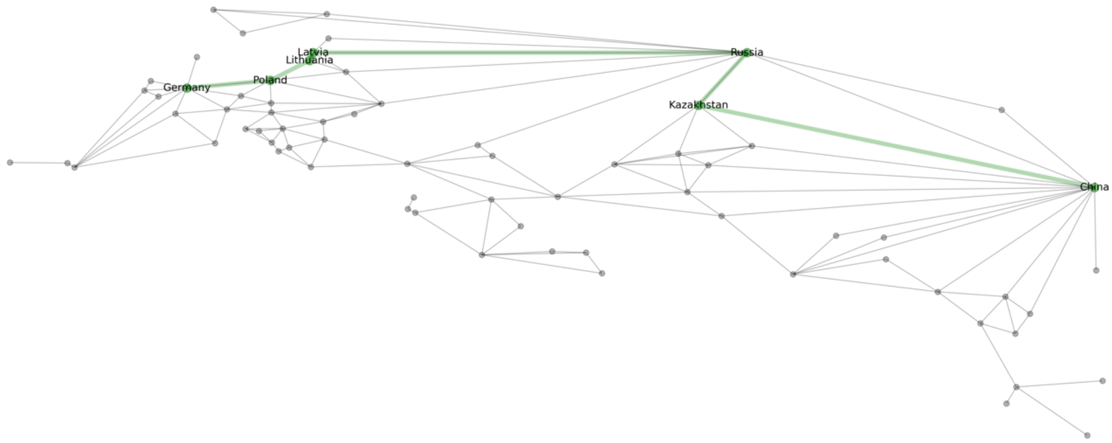
Number of edges: 4
Distance covered: 8,603 weighted kilometres

Figure 13 |
Country shortest path (D^{GL})



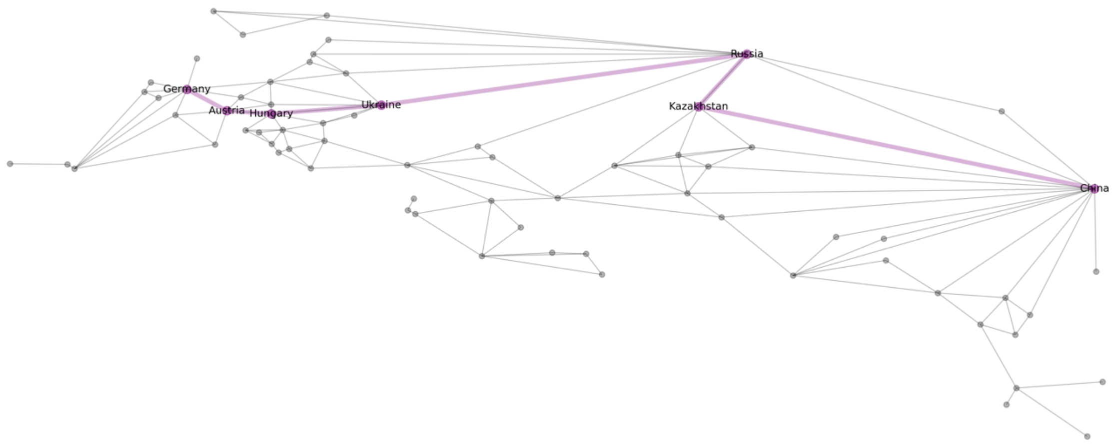
Number of edges: 4
Distance covered: 12,850 logistics adjusted kilometres

Figure 14 |
Country shortest path (D^{E})



Number of edges: 6
Distance covered: 42274 absolute difference in per capita GDP in US dollars (accumulated across edges)

Figure 15 |
Country shortest path (D^{I})



Number of edges: 6
Distance covered: 5.3 absolute institutional index distance (accumulated across edges)

Table 15 |
Results of shortest path analysis

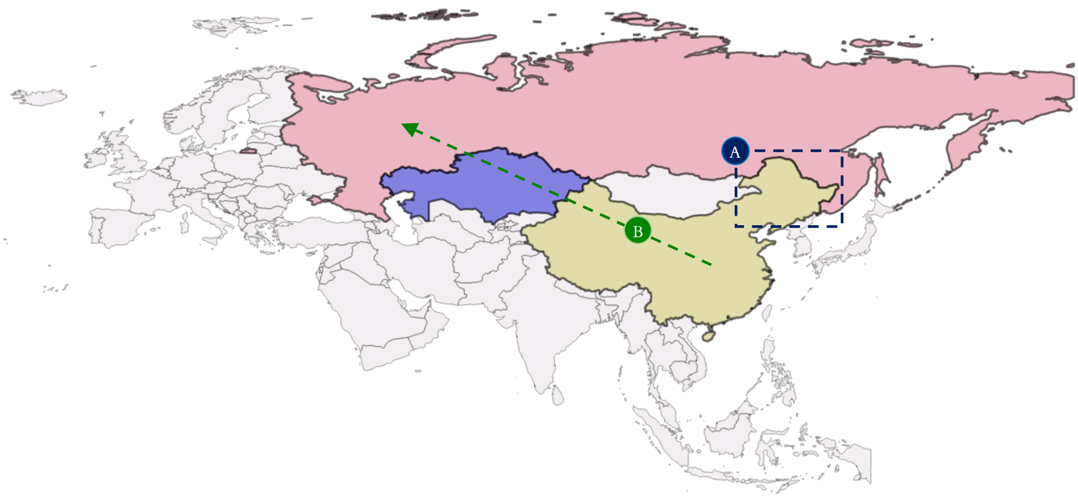
Distance	Length*	Edges	Shortest (geodesic) path
D^B	4 units	4	China => Russia => Belarus => Poland => Germany
D^{GW}	8,603 km	4	China => Russia => Belarus => Poland => Germany
D^{GL}	12,850 km	4	China => Russia => Belarus => Poland => Germany
D^E	USD 42,274	6	China => Kazakhstan => Russia =>Latvia => Lithuania => Poland => Germany
D^I	5.3 units	6	China => Kazakhstan => Russia => Ukraine => Hungary => Austria => Germany
Real reference path		5	China => Kazakhstan => Russia => Belarus => Poland => Germany

*Notes: *Units of all measures are described in greater detail alongside each figure above and Table 13*

In contrast to the 'real reference path', it is interesting to note that none of the measures managed to replicate the same series of countries. In contrast, two groups of results can be seen. The first group consists of distances D^B, D^{GW} and D^{GL} that all returned the same sequence of countries, suggesting that they involve the fewest edges and nodes which, based on these results, aligns with the least geographical distance travelled. On the one hand, the fewer edges are encouraging for they imply fewer border control checks if we assume each country checks the other's goods. Unfortunately, this is not a reliable proxy as it ignores any agreements that may facilitate the transport of goods with less friction than a typical border crossing. On the other hand, the sequence of countries and difference from our 'real reference path,' reveals an issue with geographical country level measures.

These measures, D^{GW} and by association D^{GL}, suffer from the fact that they are relying on point-to-point distances based of centroid that represent vast swathes of territory. Combined with border contiguity, this can lead to misleading results. Table 15 illustrates this below. Due to the shared border with Russia in China's north-east (area A), the data (correctly) creates an edge to which an otherwise straight-line (illustrated by B) distance is applied. This ignores the geographical reality on the ground, thus suggesting that this level of analysis whilst, directionally correct can lead to outlier results without additional variables.

Figure 16 |
Illustrating the Russia – China contiguity problem



The second group of results comes from the economic and institutional distances (D^E, D^I). Compared to the real path, these measures do come closer to approximating the path in Asia but fall short in Europe. As a result, their contribution in solving a path-finding problem may be limited. Each distance identifies the edge and pair of countries between that are the most similar either in terms of per capital income or institutional quality, these will not necessarily be strong factors in route choices, particularly for transiting goods. Additional work would be required to translate these into proxies of trade friction.

5.2.3 | Betweenness centrality

The final test assesses the betweenness centrality of each node in the network. This returns the proportion in which a node is along the shortest path between two nodes. This can be seen as a complementary measure of importance to the degrees carried out in 5.2.1. Most importantly, given the reliance on calculating shortest path possibilities, the distance measure is expected to affect the results.

Figure 17A shows the distribution of values across each distance measure for all nodes in the dataset with a generally common pattern with a power-law like distribution. This is not necessarily surprising, but demonstrating this property in this network is an important stepping stone towards corroborating other works on networks and establishing structural property which can be used in future work (Adamic et al., 2001).

Figure 17 |
[A] Country-level betweenness centrality all nodes

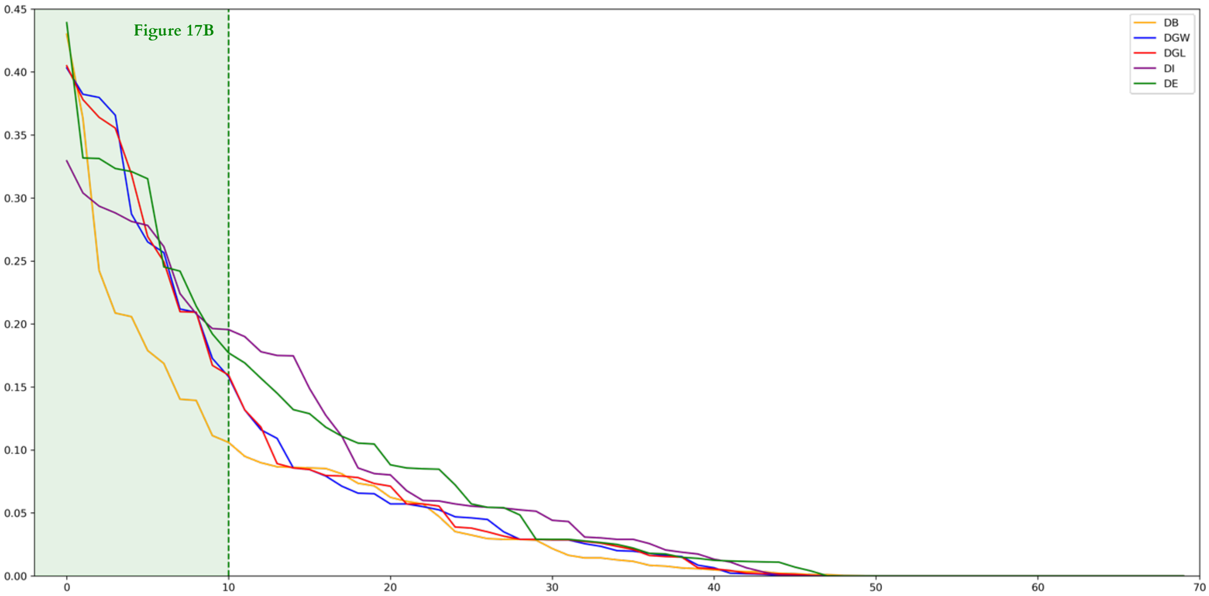
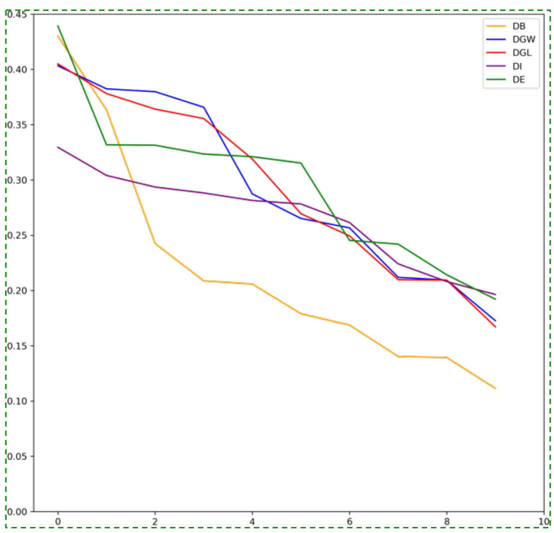


Figure 17B focuses enlarges the portion with the top 10 nodes for each measure. This clarifies the step-like nature of the value and the category-like nature of nodes present. However, with few observations it is harder to discern any further patterns without examining the nodes/countries represented.

Figure 17 |
[B]Country-level betweenness centrality all nodes



A full table of results can be found as part of the accompanying GitHub page. In line with common practice in the literature, Table 20 presents the results visualised above and with key statistical metrics in Table 16 which numerically capture the trends seen in the figures above.

Table 16 |
Summary statistics for between centrality analysis

Measure	D^B	D^{GW}	D^{GL}	D^E	D^I
Count	70	70	70	70	70
Mean	0.0514	0.0653	0.0653	0.0721	0.0750
Standard deviation	0.0837	0.1045	0.1042	0.0963	0.1055
Minimum	0.0000	0.0000	0.0000	0.0000	0.0000
25th percentile	0.0000	0.0000	0.0000	0.0000	0.0000
50th percentile	0.0121	0.0198	0.0222	0.0290	0.0235
75th percentile	0.0791	0.0698	0.0790	0.1045	0.1095
Maximum	0.4300	0.4032	0.4049	0.3295	0.4392

Notes: original table available for viewing within Notebook 210.

Looking at the constituents of each measure's top 10 in Table 20, we can observe interesting variations. The baseline distance suggests that when optimising the shortest path for the fewest edges, Russia, China and Ukraine find themselves in the top 3 and appear along the shortest route 43%, 36% and 24% respectively. This speaks to the high number of degrees they have as illustrated earlier but equally, their actual location within the network as for example, a country like Serbia does not score this highly, in fact it scores just under 9% and is in 15th place.

Table 17 |
Results of betweenness centrality analysis

Rank	D^B		D^{GW}		D^{GL}		D^E		D^I	
	Country	Value	Country	Value	Country	Value	Country	Value	Country	Value
1	Russia	0.430	Romania	0.403	Romania	0.405	Russia	0.439	Russia	0.329
2	China	0.363	Bulgaria	0.382	Turkey	0.378	Poland	0.332	Hungary	0.304
3	Ukraine	0.242	Turkey	0.380	Iran	0.364	Latvia	0.331	Iran	0.294
4	Iraq	0.209	Iran	0.366	Bulgaria	0.355	Kazakhstan	0.323	Romania	0.288
5	Turkey	0.206	Hungary	0.287	Hungary	0.319	Lithuania	0.321	China	0.281
6	Germany	0.179	Pakistan	0.265	Pakistan	0.269	Iran	0.315	Bulgaria	0.278
7	Poland	0.169	India	0.257	India	0.249	Germany	0.245	Ukraine	0.261
8	Iran	0.140	Austria	0.212	Austria	0.210	China	0.242	Turkey	0.224
9	Hungary	0.139	Iraq	0.209	Iraq	0.209	Iraq	0.214	Iraq	0.208
10	Thailand	0.111	Myanmar	0.173	Myanmar	0.167	Slovakia	0.192	Austria	0.196
…	…	…	…	…	…	…	…	…	…	…
70	-		-		-		-		-	-

Notes: table generated as part of Notebook 210 and saved in the analysis output folder, check there for full results.

Geographically determined measures (D^{GW},D^{GL}) appear almost identical, this is both in the countries in the top 10 as well as the distributions as seen from the summary statistics and the visuals. This speaks to a potentially significant methodological shortcoming and/or a limitation of using country-level information. Both relate to the attribution of the Logistics Performance Index (LPI) along a single edge.

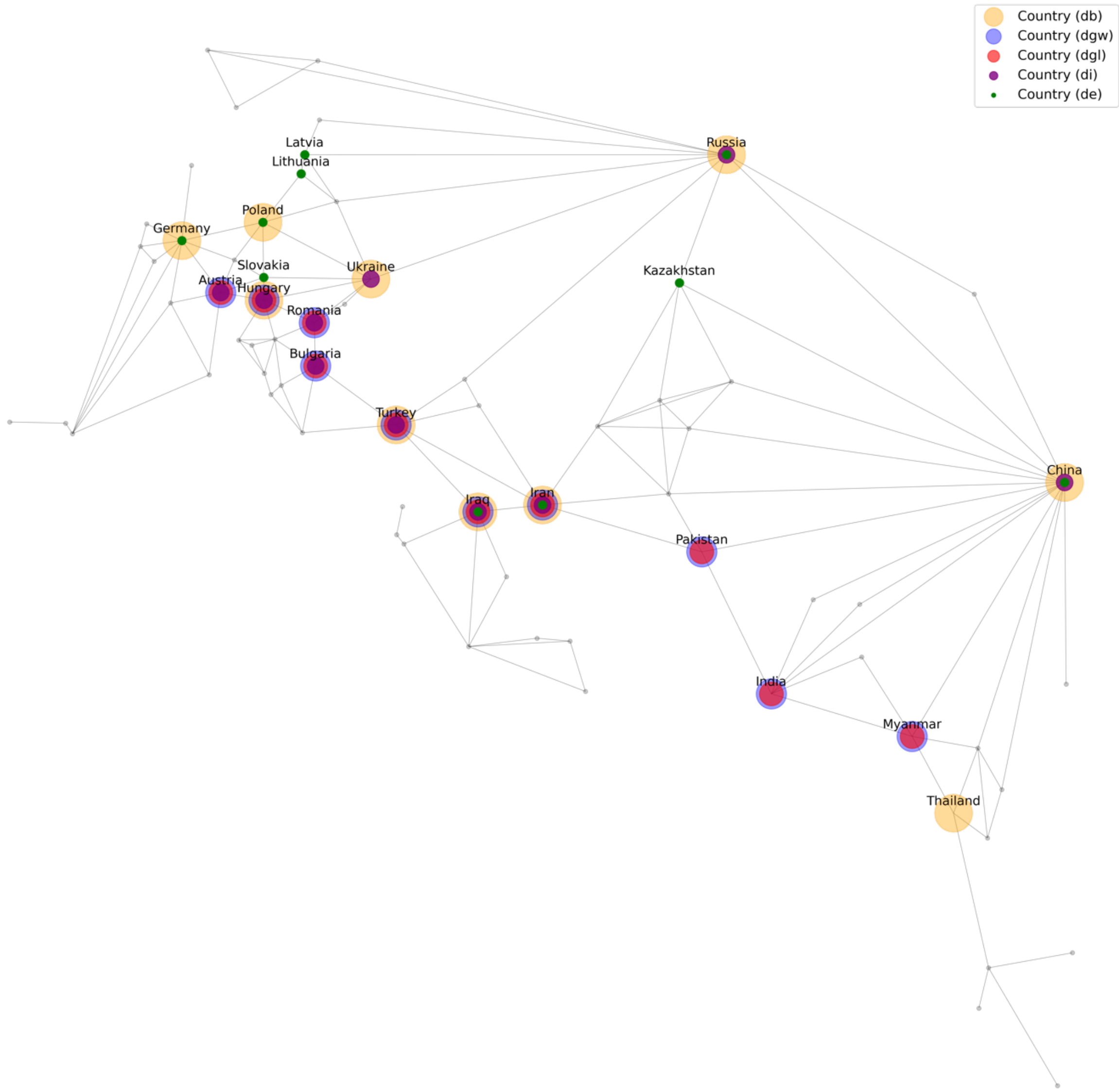
As the current methodology takes the average between two countries, it does not, nor can it, weigh each country's LPI to the length of infrastructure within said country. For instance, if country A and country B are 100 kilometres apart according to $Data_{grav}$ with each country having an LPI of 2 and 5 respectively, the calculated DGL would be 165 km.[10] However, a given path or even the landmass of a state may mean that in practice goods travel 30km on LPI 2 infrastructure and 70 at LPI 5 which would give us 122.5, an almost 25% over the originally calculated distance. This speaks particularly strongly to the potential advantages of using route-level data, the section below will investigate this.

[10] [100 * (1+((2+5/2)-1)/4)] taking the formula described in the methodology section. $D_{ij}^{GL} = \frac{\text{LPI}_{pi} + \text{LPI}_{pj}}{2} \times distance_{ij}$

Finally, D^E and D^I yield different result to each other and the remaining measures, nevertheless are fall to the same critique as they did in the standalone shortest path test.

In closing of this test and the country-level networks, Figure 23 captures the results of the betweenness test, in the network itself. The visualisation highlights nodes that appear in the top 10 across the distances. Countries such as Iraq and Iran are the only to appear across all measures, with Turkey and Hungary appearing in all except D^E. Finally, Russia and China appear across three measures, although notably neither of the geographically determined distances.

Figure 18 |
Country-level betweenness centrality key nodes visualised in network



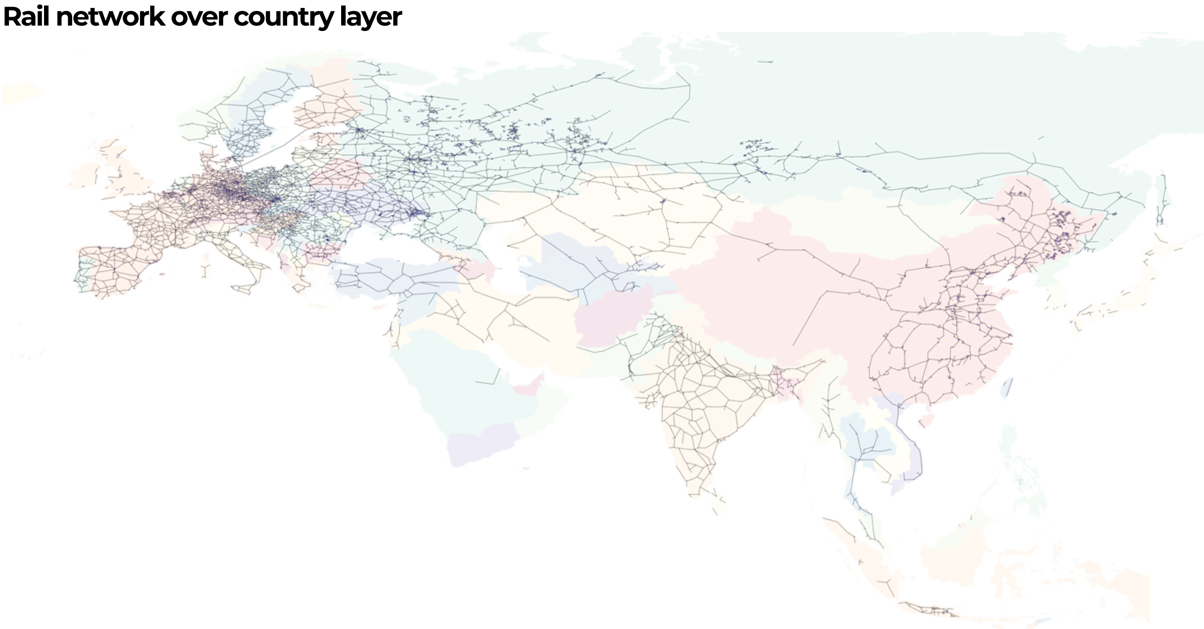
5.3 | Route-level networks

What are the findings for route-level networks?

For a consistent comparison, the route-level networks are constructed with the same geographical scope as the preceding analysis in Section 5.2. The results presented in this section are all contained in Notebook 220, which takes the merged dataset [Doc125] alongside route level data, representing rail connections across Europe and Asia from $Data_{NE}$. The below figures showcase the networks, with Figure 19 outline the routes and nodes while Figure 20 that adds country boundaries for context. Table 18 provides the summary statistics for $Data_{NE}$ below.

Figure 19 |
Rail network only visualisation

Figure 20 |
Rail network over country layer



The main elements taken from $Data_{NE}$ are categorical variables outlining to which country (iso3) code a rail segment belongs to and the unadjusted geographical length of each segment in kilometres (D^G) and the same length adjusted by the logistics factor (D^{GL}). From the numbers we can see that the majority of links fall into the 0 – 29 or 40.5 km range for D^G and D^{GL} respectively with a rapid increase in edge length after that. This would be expected to be a problem were distances and paths calculated purely on edges, however the use of geographical distance mitigates this shortcoming.

Table 18 |
Summary statistics for in-scope datapoints for $Data_{NE}$

Measure	Countries	Rail segment ID	D^G (km)	D^{GL} (km)
Data frame column name	ADM0_A3	rwdb_rr_id	length (raw data in metres)	length_lpi (raw data in metres)
Count	70	20528	20528	20528
Mean	–	–	26.3	36.8
Standard deviation	–	–	40.8	57.2
Minimum	–	–	0	0.1
25th percentile	–	–	5.2	7.5
50th percentile	–	–	12.1	17.3
75th percentile	–	–	29.0	40.5
Maximum	–	–	1038.3	1385.8

Notes: original table available for viewing within Notebook 220.

Nevertheless, it is interesting to understand the network from the number of edges per country. Once more, a power-law like distribution can be observed with Russia, China, Germany alone accounting for 50% of rail links and the first 9 countries accounting for approximately 80% (as shown by the reference lines in Figure 23). This can be seen consistently across the two other measures D^{BG} and D^{GL} in Figure 24.

Figure 24 presents the amount of rail kilometres (unadjusted and LPI adjusted - D^{DG} and D^{DGL}) along with the cumulative summary. It is expected that the ranking of countries would be different when using geographically meaningful measures compared to the number of edges (links). In line with this expectation, while the countries contributing the most to any of the three measures remains comparable, their order changes, sometimes significantly. A demonstration of this phenomenon are India and Germany who shift apart by 2 positions from one measure to the next. This speaks to the relatively poorer quality of infrastructure in India. Similarly, the narrowing of the gap between Poland, Ukraine and Germany also speaks to their comparatively lower quality infrastructure index which results in a higher kilometre multiplier.

This is an important illustration of the role that an accurate reflection of distance and by implication trade friction along routes would have on identification of key nodes, shortest paths and ultimately the flow of goods along the network.

Finally, the summarised data speaks to the extent to which cross border edges are present in the dataset. These represent two limitations of the data in its current form. Firstly, as the line itself did not fall exactly within the border of any of state, it is not assigned to any country code. As a result, it is a distorting factor to any country level aggregations. Secondly, as a result of the first limitation, no adjustment could be made as per the logistics performance index. As a result, the geographical distance was left as is with an assumed LPI of 1. In alternative scenarios, a very hard or poor border infrastructure could be approximated by raising this factor significantly compared to other links. However, this is not pursued in this work.

Figure 21 |
Rail route per country + cumulative percentage (D^B)

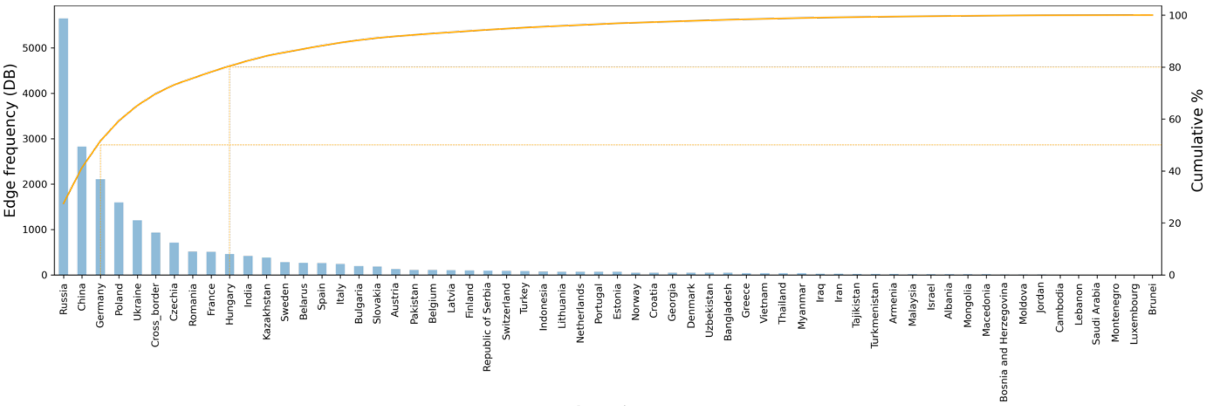
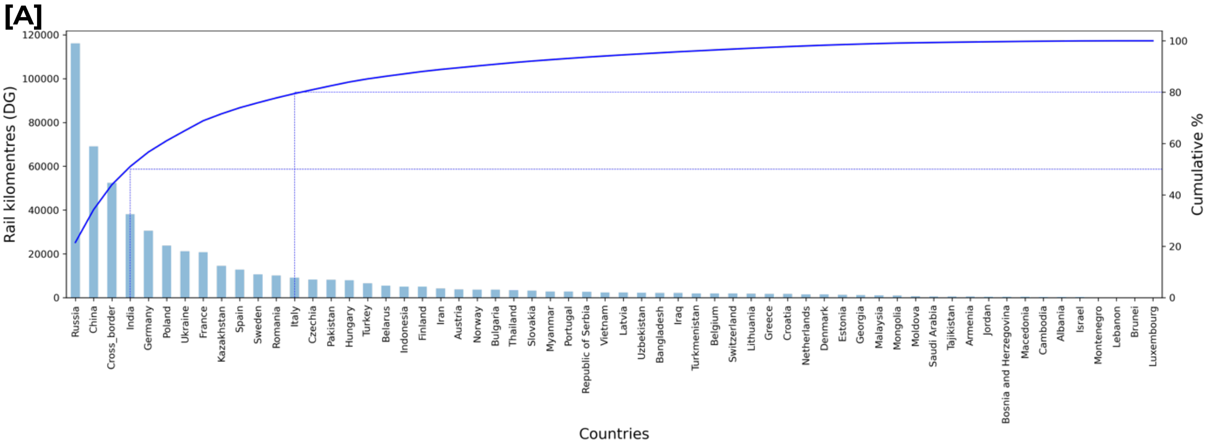
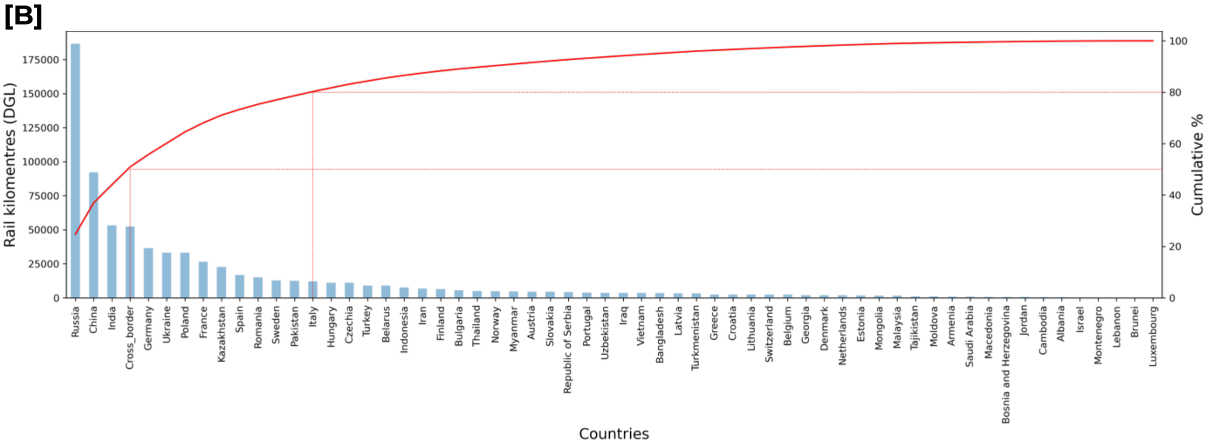


Figure 22 |
Rail distance per country + cumulative percentage (D^{DG} *and* D^{DGL})

[A]



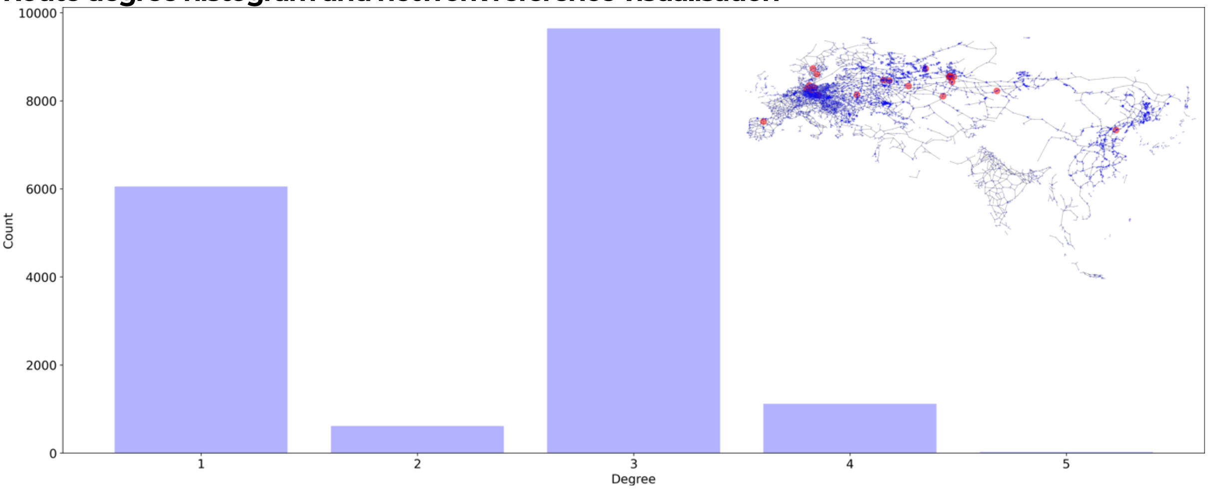
[B]



5.3.1 | Measuring network degrees

Assessing the degrees of each node in this network comprised of physical infrastructure we can observe key characteristics in Figure 23. Firstly, the range of the degrees (1-5) continues to be in line with observations by Cardillo et al. (2006). This speaks to the properties of this network, which although on a bigger geographical scale, given the extent of connectivity could behave in line with the smaller scale networks studied by Cardillo et al. and thus lend further avenues to explore in terms of analysis.

Figure 23 |
Route degree histogram and network reference visualisation



Furthermore, there is additional insight by way of where key nodes are. Although visualised, Table 19 summarises the occurrence of 5-degree and 4-degree nodes per country.

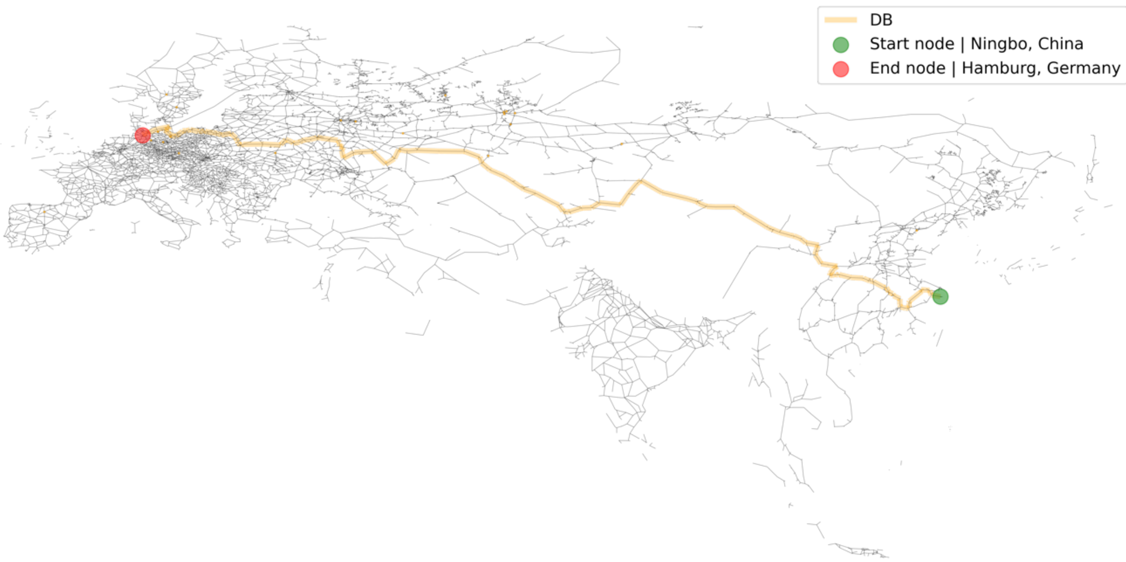
Table 19 |
Results of betweenness centrality analysis

Country	Nodes with 5 edges	Nodes with 4 edges	Total (total rank)
Russia	9	272	281 (1)
Germany	4	133	137 (3)
Kazakhstan	2	21	23 (7)
Sweden	2	19	21 (8)
Ukraine	1	92	93 (5)
China	1	106	107 (4)
Czech Republic	1	39	40 (6)
Spain	1	18	19 (9)
Poland	-	144	144 (2)
Others	-	272	272 (-)
Total	21	1116	1137

5.3.2 | Shortest path analysis

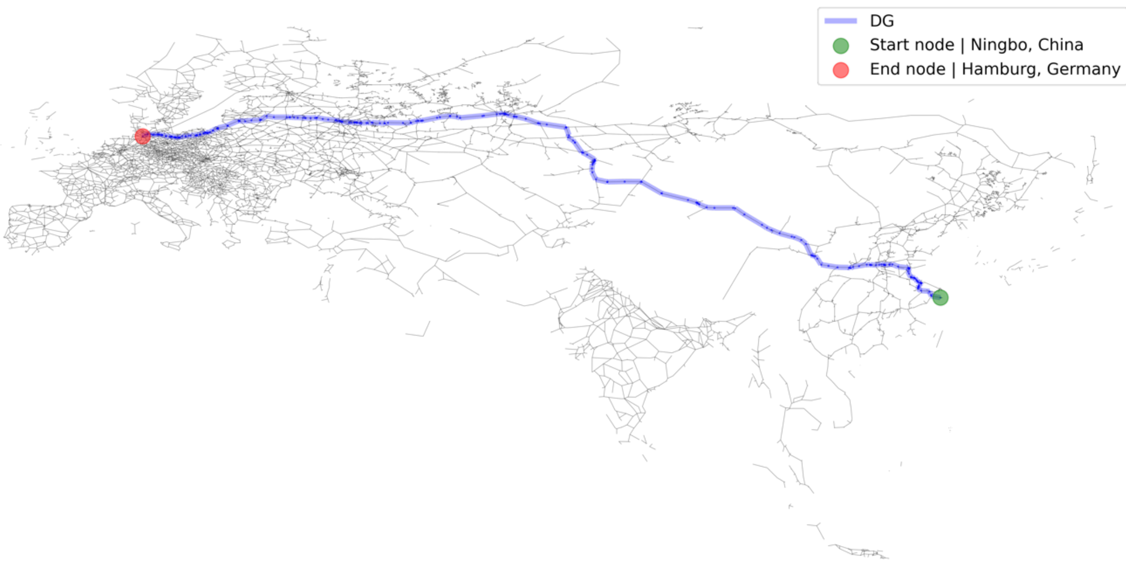
The shortest path analysis at route level differs in two important ways in comparison to the country-level assessment. Firstly, there are fewer measures of distance that can be studied. Specifically, as discussed previously D^E and D^I are challenging to translate down to routes without them becoming a proxy for the number of edges covered along a particular path. As a result, only the baseline and geographical measures remain and are useful to carry out a comparison. Secondly, the origin and destinations are to an extent arbitrary, and can be amended to run a number of experiments through two lines of code in the accompanying material. In this case, we are still comparing a route from China to Germany and choosing Ningbo and Hamburg as the respective points within each country (Wang et al., 2019). Both are important industrial and transport hubs and situated at the extreme end of each country. Identifying the shortest path using the code in Notebook 220 results in the following results presented as part of Figure 24 through to Figure 26 individually and overlaid in Figure 27 with quantified distances in Table 19.

Figure 24 |
Route shortest-path route (D^B)



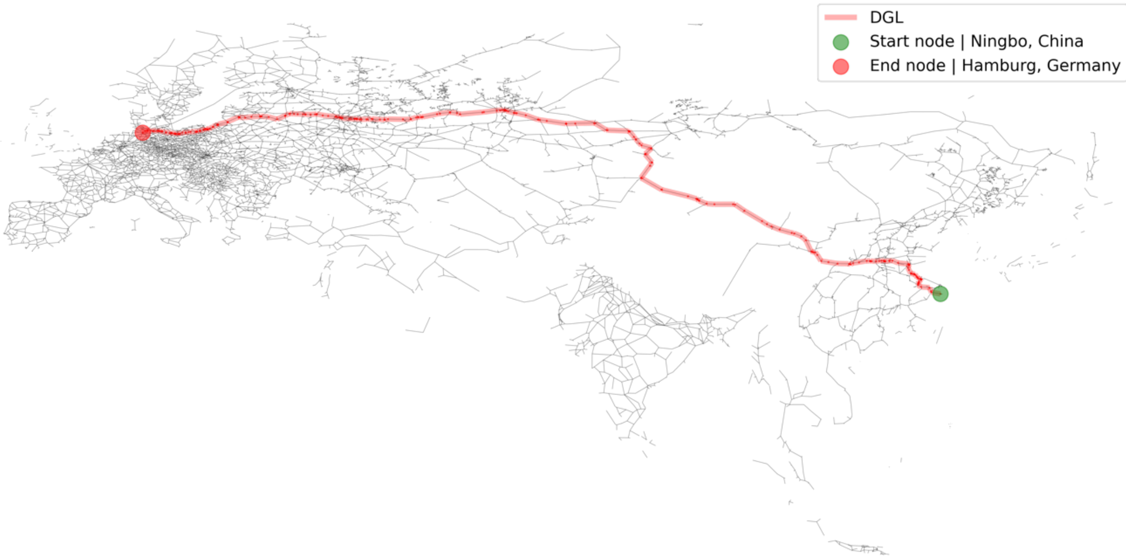
Number of edges: 202
Distance covered: 202 edges

Figure 25 |
Route shortest-path route (D^G)



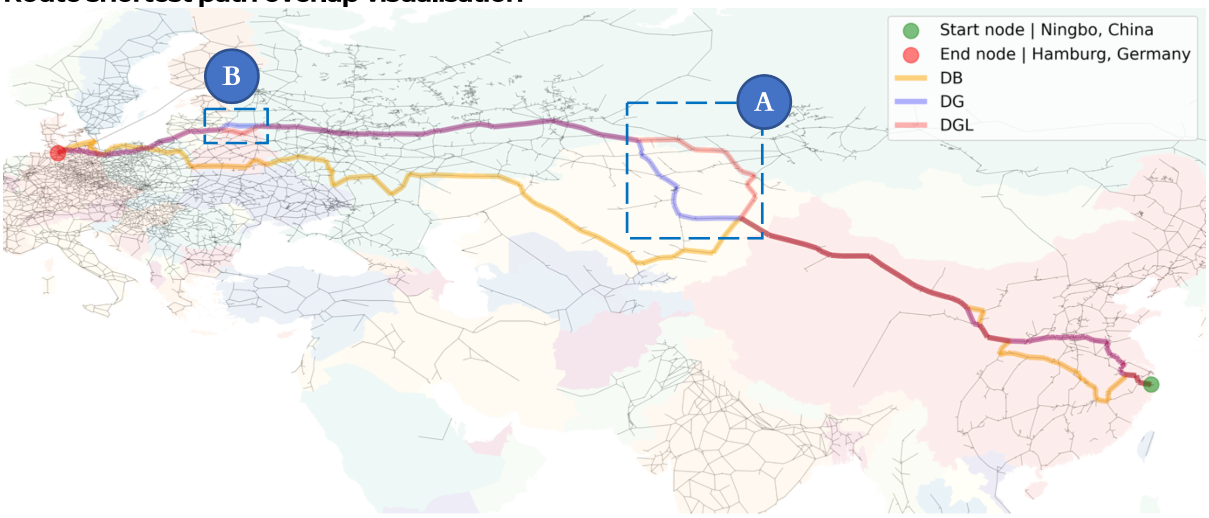
Number of edges: 276
Distance covered: 10,844 kilometres

Figure 26 |
Route shortest-path route (D^{GL})



Number of edges: 276
Distance covered: 15,304 logistics adjusted kilometres

Figure 27 |
Route shortest path overlap visualisation



As previously, any disparity between D^B and the others is expected, in this instance Figure 27 shows the choice of longer (and therefore fewer) edges for D^B in comparison to D^G and D^{GL}. What is interesting between the latter two measure is their disparities over Kazakhstan (A) and the Belarus / Latvia border (B). In both instances, the quality of infrastructure has a large enough effect to alter the course of the route. In the case of (A) the inference can be that given the higher penalty due to the LPI adjusted distance, it is most efficient to find the shortest path to a country with a better quality of infrastructure, in this case Russia. Similarly, in (B) the choice to go into Belarus and lessen the distance in Russia there is a potentially attractive route/corridor provided other not accounted for frictions such as border crossings do not offset this. Table 20 below quantifies the results of each path below with emphasis on a side-by-side comparison of D^G and D^{GL}.

Table 20 |
Results of route-level shortest path analysis

D^G results based on path from geographical (non-adjusted) distance

Country	Kilometres	LPI km	km/edge	lpi km/edge
Data column	length	length_lpi	calculated	calculated
China	4685	6253	54	72
Russia	2767	4448	33	53
Kazakhstan	1871	2923	62	97
Poland	441	615	12	16
Germany	328	391	15	18
Lithuania	252	338	50	68
Latvia	93	131	31	44
Cross border*	412	412	59	59
Total	10849**	15512	39	56

D^{GL} results based on path from logistics adjusted distance

Country	Kilometres	LPI km	km/edge	lpi km/edge
Data column	length	length_lpi	calculated	calculated
China	4685	6253	54	72
Russia	3250	5223	33	53
Kazakhstan	859	1342	78	122
Poland	441	615	12	16
Germany	328	391	15	18
Lithuania	252	338	50	68
Belarus	87	144	87	144
Latvia	18	26	6	9
Cross border*	970	970	88	88
Total	10891	15304	39	55

D^B results based on path from edge = 1 distance

Country	Kilometres	LPI km	km/edge	lpi km/edge
Data column	length	length_lpi	calculated	calculated
China	5468	7299	76	101
Kazakhstan	2856	4461	124	194
Russia	1930	3102	58	94
Poland	834	1162	23	31
Belarus	638	1053	53	88
Germany	229	273	19	23
Cross border*	1277	1277	98	98
Total	13232	18627	66	92

*Notes: * cross border entry represents edges that go across borders and were not able to be neatly allocated to a country on either side*
***Total is higher by 5km due to rounding.*

The highlighted areas of the table above, showcase areas (A) and (B) from the map on the previous page. It is notable that while in terms of kilometres, the D^{GL} route is longer by 42km, comparing LPI, the D^{GL} is shorter by 208 LPI adjusted kilometres or a factor of almost 5. However, although the variance in country logistics quality is a factor, the element of 'cross border' edges is proving troublesome. Once again, this is an issue due to assumed LPI index of 1 which does not affect the distance covered in comparison to surrounding edges, therefore, the route-finding algorithm is encouraged to border hop. This suggests that more time needs to be spent curating the data and re-running the analysis in Notebook 220.

5.3.3 | Betweenness centrality

This analysis approach is consistent with the approach taken for countries in Section 5.2.3. Cycling through the three available measures, betweenness centrality is calculated and presented below. Figure 28 and its subplots visualise the distribution of results while Table 21 contains the relevant summary statistics.

Figure 28 |
[A] Route-level betweenness centrality all nodes

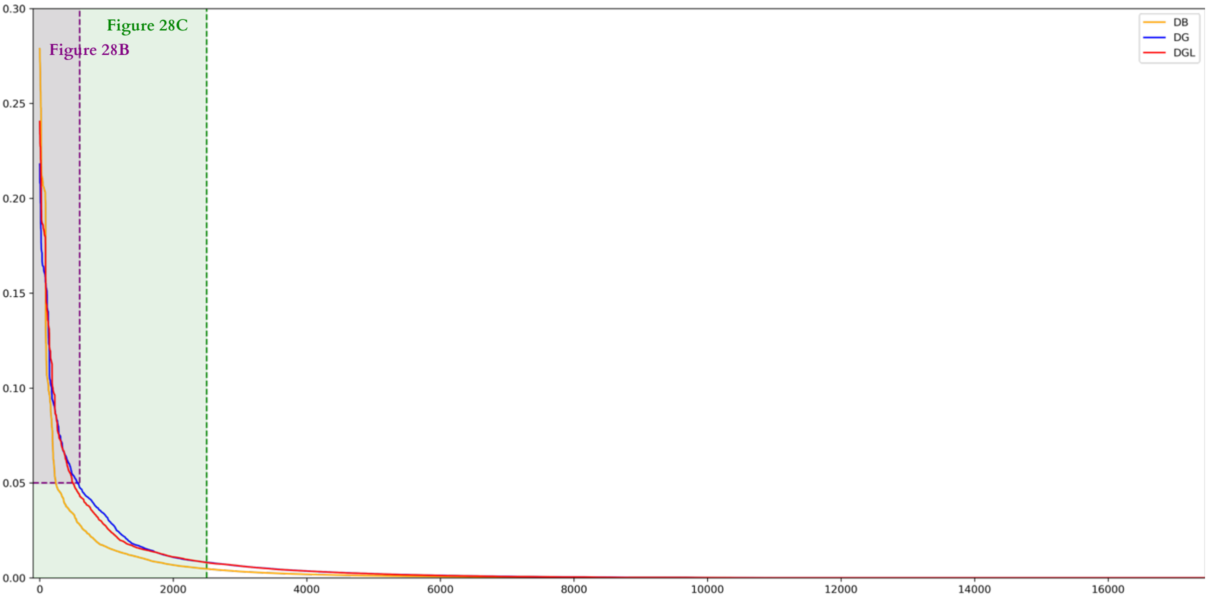


Figure 28B | Node betweenness centrality (nodes greater than 0.05)

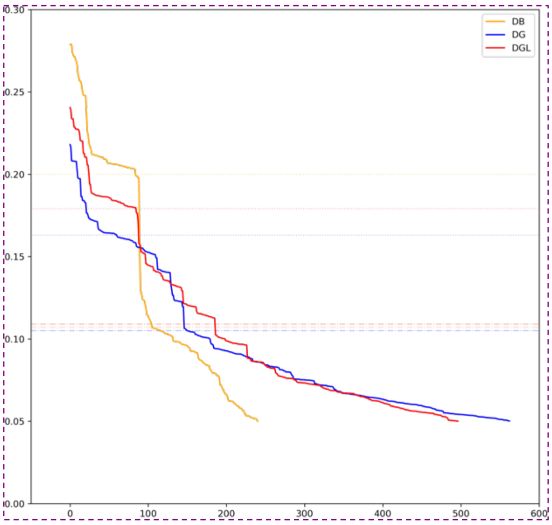
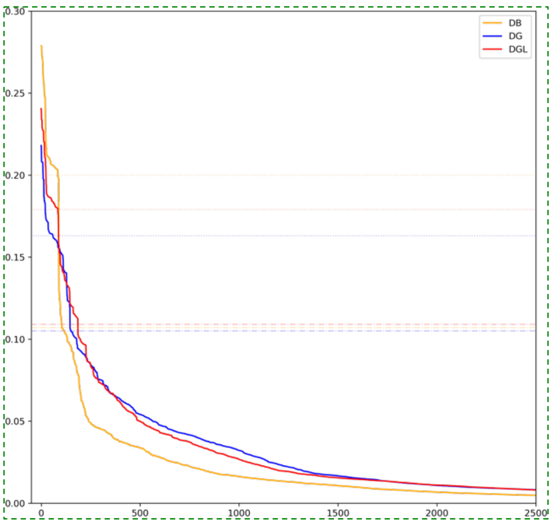


Figure 28C | Node betweenness centrality (first 2500 nodes).



The results across all measures follow the now-expected power-law distribution pattern. As a result of more datapoints, a more prominent 'step-like' nature is evident as Figure 28B demonstrates. The results are further broken down in the table below.

Table 21 |
Summary statistics betweenness centrality analysis of *Data*$_{NE}$

Measure	D^B	D^G	D^{GL}
Count	17449	17449	17449
Mean	0.0044	0.0062	0.0061
Standard deviation	0.0187	0.0194	0.0202
Minimum	0.0000	0.0000	0.0000
25th percentile	0.0000	0.0000	0.0000
50th percentile	0.0002	0.0002	0.0002
75th percentile	0.0015	0.0029	0.0030
Maximum	0.2788	0.2180	0.2405

Notes: original table available for viewing within Notebook 220.

In addition to numerically representing the distribution of the values in the graphs above, the summary statistics suggest a significant decrease of observations by just over 3000. This can mostly be explained by a large number of single nodes remaining as remnants of filtering the data to the desired continents and eliminating in keeping with the in-scope countries.

As before, it is important to now consider *where* these nodes are. In contrast to previous analysis, we can take advantage of the step-like nature of the node distribution to approximate three categories of node per each measure. These are the nodes between the reference lines in Figure 28B for all the nodes with betweenness centrality values greater than 0.05. The categories follow the same rule where Category 1 is always from the maximum value to the top-most reference line, Category 2 between two reference lines of the same colour and Category 3 from the bottom-most reference line of the given measure to the minimum value of 0.05.

The figures below, present the nodes for their respective measure but also highlight those in common with other measures by category type. This provides additional information to us about the network as a whole as a result of the fact that those specific nodes are consistently highlighted regardless of the measure. This speaks to the potential utilisation along shortest paths in the network but equally, the susceptibility of the network to the node being taken out of the network. In practical terms this could be a fault or gradual worsening of quality that affects the way transport passes through. These kinds of enquiries examine the resilience of the network and could be enabled by such dataset and construct in future work.

Figure 32 summarises the preceding figures in one visualisation to provide an alternative snapshot of which nodes commonly occur together with a betweenness centrality score greater than 0.05 as well as the outline of the respective countries for greater geographical context.

Figure 29 |
Route-level betweenness centrality key nodes visualised (D^B)

Figure 30 |
Route-level betweenness centrality key nodes visualised (D^G)

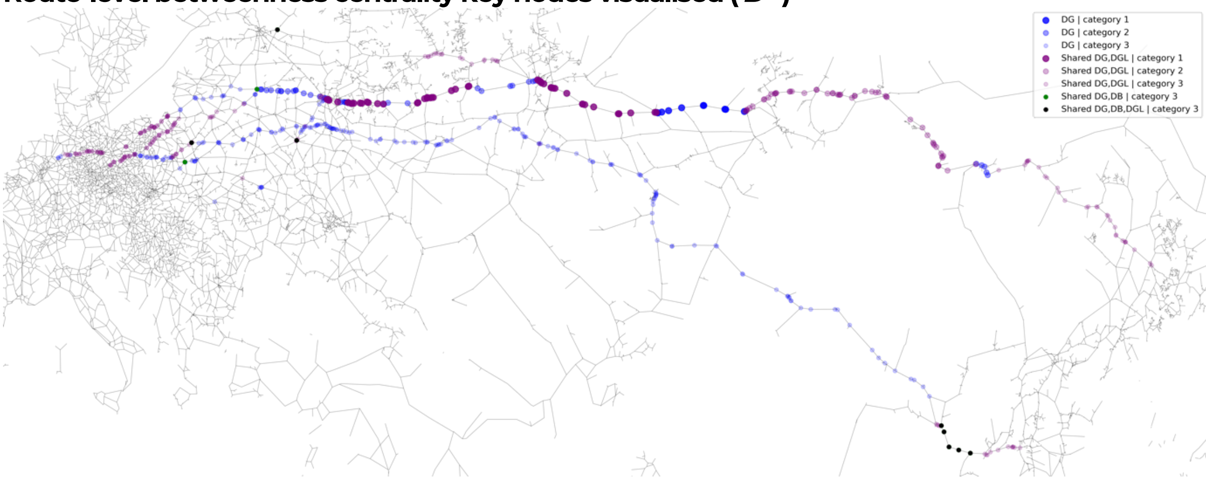


Figure 31 |
Route-level betweenness centrality key nodes visualised (D^{GL})

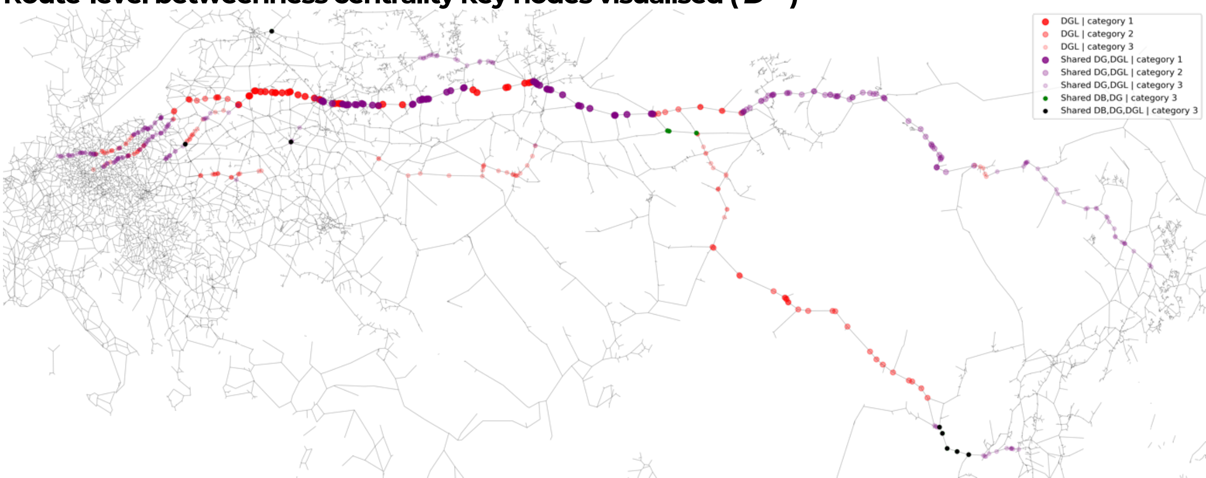


Figure 32 |
Route-level betweenness centrality key nodes visualised (layered)

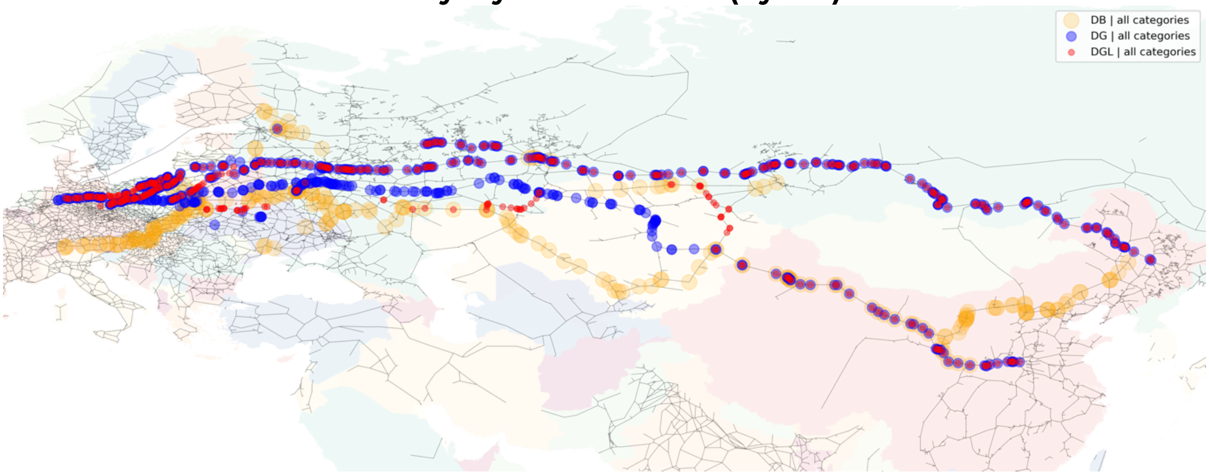


Table 22 provides a summary of the countries in which these nodes can be found. Russia, China and Poland are consistently found in the top 3 countries as measured by the number of nodes in the top 500 nodes of the given results. The 500 mark was selected as approximately the equivalent position along the distribution as the country-level analysis.

One of the main conclusions to be drawn from the visualisation and the summary table below is the nodes that are likely to be key in the long-distance transport of traded goods. An extension of this property could be utilised to run scenarios and add impacts of specific projects along BRI routes to either see their effect on pre-existing key nodes or whether project of significant enough scope can create new short routes/corridors.

Table 22 |
Results of betweenness centrality analysis (first 500 nodes)

Rank	D^B		D^G		D^{GL}	
	Country	Node count	Country	Node count	Country	Node count
1	Russia	197	Russia	263	Russia	230
2	China	85	Poland	80	Poland	107
3	Poland	50	China	57	China	58
4	Kazakhstan	28	Kazakhstan	37	Germany	37
5	Ukraine	26	Belarus	19	Ukraine	19
6	Sweden	19	Germany	19	Cross border*	13
7	Belarus	19	Lithuania	9	Lithuania	12
8	Slovakia	18	Cross border*	6	Belarus	10
9	Cross border*	18	Ukraine	6	Kazakhstan	10
10	Austria	15	Latvia	4	Latvia	4
11-17	Others	25	-	-	-	-

Notes: table generated as part of Notebook 220 and saved in the analysis output folder, check there for full results.

6 | Discussion, further work and concluding comments

6.1 | Overview

What does this section cover?

Section 5 contains details analysis of country and route-level networks in isolation, this section builds on its findings to comment on the contrast and similarities of the findings and reflect on the contribution this makes to the specific research questions and the implication on the overarching research agenda.

6.2 | Comparison of analysis results

How comparable are country and route level results?

Assessing the degree distribution of each network primarily served to provide awareness about the network and properties that could be inferred. The country-level network covered a wider range from 1 – 14 with countries with large territories naturally bordering with more countries and hence enjoying a higher degree count. In contrast, the route-level analysis resembled smaller scale infrastructure networks such as city streets with its degree range of 1-5 (Cardillo et al., 2006). Neither is superior to the other, however, in the case of route-level networks, a wider body of work can be leveraged in studying properties from the ground up.

A test with markedly differing results is the shortest-path test. Across both networks a path is sought from China to Germany using an assortment of distance measures used in the literature of international trade alongside a baseline and adjusted measure incorporating infrastructure quality with summary results shown in Table 23.

Table 23 |
Country-level and route-level shortest-path result comparison

Measure (units*)	Country-level analysis	Route-level analysis	Difference (%)
D^B (edges)	4	202	+198 (not relevant)
D^G (km)	-	10,844	+2,241 (+26%)
D^{GW} (km)	8,603	-	
D^{GL} (km*)	12,850	15,504	+2,654 (+21%)
D^E (USD)	USD 42,274	-	-
D^I (units)	5.3 units	-	-

*Notes: *Units of all measures are described in greater detail alongside each figure above and Table 13 ** kilometres adjusted by Logistics Performance Index (LPI)*

Analysis	D^G / D^{GW} path	D^{GW} path
Country-level	China => Russia => Belarus => Poland => Germany	China => Russia => Belarus => Poland => Germany
Route-level	China => Kazakhstan => Russia => Latvia => Lithuania => Poland => Germany	China => Kazakhstan => Russia => Belarus => Latvia => Lithuania => Poland => Germany
Difference	+ Kazakhstan + Latvia + Lithuania - Belarus	+ Kazakhstan + Latvia + Lithuania

Among directly comparable measures, an increase in the geographical distance of more than 20% is observed across both distance measure pairs. This speaks to country to country distance values that are removed from the on-the-ground reality yet are at the heart of international trade analysis.

Secondly, the shortest path-finding exercise revealed a compounding effect of border contiguity and country to country distances that do not reflect the actual path taken. In both country-level analyses the shortest path involves a move directly from China to Russia in contrast to the route-level path of going through Kazakhstan (also a reflection of the state of real transportation today). The shortcoming is of correctly modelling or capturing phenomena such as internal distance travelled within a country to reach the next node along the network. On this front, the route-level analysis showcases that there is significant value in bottom-up analysis and corroborates calls in both transport and international trade literature for the incorporation of transport and transportation routes (Baier et al., 2018).

Finally, comparing the betweenness centrality results shows common properties across both networks when it comes to their power-law like distribution. Once again, this is a useful trait to be aware of when examining the network as a whole. However, considering the lists of countries, in particular for the directly comparable geographically based distance measures, few commonalities present themselves. This suggests that countries most significant by way of key nodes in a wider network are fewer and in line with the Eurasian land bridge.

6.3 | Answering the research questions

What is the implication of results on the research questions?

Taking each research question in turn, RQ3 enquired about the impact of distance as a means of informing the answers and further work that would contribute to answering RQ2 and RQ1. The comparison above, along with the standalone analysis suggests that abstract measures such as D^E and D^I while serving an important role in understanding proximity of countries in a multi-dimension space, do not provide particularly revelatory insight to path selection and therefore did not contribute to understanding the BRI from this perspective. However, varieties of geographical distance, beyond pure geographical distance shows signs of promise albeit more works remains to be done. Most notable is the use of bottom-up distance estimates to generate physical distance measures that are more reflective of the world. In this work, this has led to the identification of routes that follow real-world patterns but also, when an adjustment for the quality of infrastructure was included, identification of new routes through the avoidance of perceivably lower connections.

Although the impact on the flow of goods has not been assessed, the construction of the network, the respective datasets and the data processing workflow provide a grounding on which such additional work could be carried out.

Building up to RQ2, the additional information that network based approaches (which could ultimately serve in the construction of a wider system) is largely dependent on the resolution of the data utilised in the construction of the network. A country-level network would yield some but limited *extra* information given the prominence of countries as the basic unit of analysis. A route-level construct provides additional information through the benefits outlined as part of the above, but further the potential identification of alternative corridors through the inclusion of additional variables to construct a more robust representation of the state of infrastructure and connectivity along the BRI.

6.4 | Limitations

What are the limitations of the scope and approaches?

Limitations in this project can assessed as limitations of the proof of concept itself and limitations of the methodologies used. In the case of the former, limitations were acknowledged from the start of the work, assumptions / constraints limiting the geographical scope but perhaps more importantly limiting the mode of transport to rail only was not expected to lead to results that are generally applicable in trade. This is particularly the case in the identification of shortest routes, where the exclusion of sea but also road connections limits the connections that can be had. For instance, alternative short paths cannot be found if naval links across the black sea or even between Turkey and Greece across the Bosporus are not in place.

This sheds light on the latter limitation with the data used for the analysis in this work. Limitations revealed themselves with the allocation of rail routes (network edges) to countries where ~5% of connections were categorised as cross-border. This factor can be overcome with better management of the $Data_{NE}$ and the processing in the QGIS process block and Notebook 220. The ability to treat each data block in a modular fashion is in contrast a strength of this project and the modularity built into the coding component as per Figure 4 and Figure 7 and satisfies the objective set out in Table 6.

6.5 | Further work

What further avenues can be explored?

Further work can focus on refinement within a specific model or on developing or incorporating parallel models that would interact with a route-based network. Refinement within this model takes its cues from the limitations above and comes down to incorporation of additional data sources to better reflect existing connectivity. There are two further extensions to this:

1. The incorporation of BRI projects and other factors that may affect either distance or trade friction (as two significant determinants of international trade), this would provide the space for simulations and assessment of impact of a project on the network or system as a whole.

2. The incorporation of spatial interaction models to observe the *flow* of traded goods. Once more, the inclusion of specific routes / corridors is expected to provide more interesting insight compared to the currently dominant country to country focused research agendas.

Interaction between models would centre on inputs and outputs between them. These could be directly complimentary models around congestion of routes, micro-optimisation of freight rail scheduling to macro components such as economic growth. The latter would be an interesting component to incorporate as it would provide the final factor in the baseline gravity equation (economic size, distance, trade friction). These avenues follow the nesting and positioning of systems and sub-systems in Figure 1.

6.6 | Conclusion

This project set out to explore means of modelling distance and infrastructure as a means of assessing the utility of route-level networks in a process towards building systems as a means of understanding and modelling international trade. The Belt and Road Initiative served as both a catalyst for the relevance of such an endeavour and the ultimate focus of its proof of concept applicability. While limitations are highlighted and acknowledged, it has been shown that route-level approaches do provide additional insight compared to traditional approaches. The contribution of this work to the overarching research question of *how* a construct such as the BRI could be modelled stems from both the results showing the promise of networks as well as the accompanying data frameworks and code published along-side it.

7 | Bibliography

Adamic, L.A., Lukose, R.M., Puniyani, A.R., Huberman, B.A., 2001. Search in power-law networks. Physical review E 64, 046135.

Anastasiadou, I. (E), 2019. Iron Silk Roads: the geopolitics of past and present initiatives for the revival of Eurasian trade through overland transport corridors. Cambridge Journal of Regions, Economy and Society 12, 57–75. https://doi.org/10.1093/cjres/rsy035

Baier, S.L., Kerr, A., Yotov, Y.V., 2018. Gravity, distance, and international trade, in: Handbook of International Trade and Transportation. Edward Elgar Publishing.

Baldwin, R.E., Kay, D.A., 1975. International trade and international relations. International Organization 29, 99–131.

Barbieri, K., Keshk, O., 2016. Correlates of war project trade data set codebook, version 4.0. Online: http://correlatesofwar. org.

Barrett, P., 2005. Euclidean distance: Raw, normalised, and double-scaled coefficients. The Technical Whitepaper Series 6, 1–26.

Belt and Road Portal, 2019. 已同中国签订共建"一带一路"合作文件的国家一览 (List of countries that have signed cooperation documents with China to jointly build the "Belt and Road" - via google translate) [WWW Document]. Belt and Road portal, yidaiyilu.gov.cn. URL https://www.yidaiyilu.gov.cn/gbjg/gbgk/77073.htm (accessed 8.3.20).

Berthelon, M., Freund, C., 2008. On the conservation of distance in international trade. Journal of International Economics 75, 310–320.

Bērziņa-Čerenkova, U.A., 2016. BRI Instead of OBOR – China Edits the English Name of its Most Ambitious International Project [WWW Document]. Latvian Institute of International Affairs. URL https://liia.lv/en/opinions/bri-instead-of-obor-china-edits-the-english-name-of-its-most-ambitious-international-project-532 (accessed 6.28.20).

Blonigen, B.A., Wilson, W.W., 2018. Handbook of International Trade and Transportation. Edward Elgar Publishing.

Bokhari, F., Findlay, S., Yu, S., 2020. Pakistan seeks relief from China over Belt and Road [WWW Document]. The Financial Times. URL https://www.ft.com/content/4af8101b-599c-407d-8850-3fd27cd9b31c (accessed 7.28.20).

Borchert, I., Yotov, Y.V., 2017. Distance, globalization, and international trade. Economics letters 153, 32–38.

Brakman, S., Frankopan, P., Garretsen, H., Van Marrewijk, C., 2019. The New Silk Roads: an introduction to China's Belt and Road Initiative. Cambridge Journal of Regions, Economy and Society 12, 3–16. https://doi.org/10.1093/cjres/rsy037

Brandes, U., 2001. A faster algorithm for betweenness centrality. Journal of mathematical sociology 25, 163–177.

Buckley, P.J., 2020. China's belt and road initiative: Changing the rules of globalization.

Cardillo, A., Scellato, S., Latora, V., Porta, S., 2006. Structural properties of planar graphs of urban street patterns. Physical Review E 73, 066107.

Chan, M.H.T., 2018. The Belt and Road Initiative–the New Silk Road: a research agenda. Journal of Contemporary East Asia Studies 7, 104–123.

Chaney, T., 2018. The gravity equation in international trade: An explanation. Journal of Political Economy 126, 150–177.

Costinot, A., 2009. On the origins of comparative advantage. Journal of International Economics 77, 255–264. https://doi.org/10.1016/j.jinteco.2009.01.007

De Soyres, F., Mulabdic, A., Murray, S., Rocha, N., Ruta, M., 2018. How much will the Belt and Road Initiative reduce trade costs? The World Bank.

Diamond, J.M., 2005. Guns, germs, and steel : the fates of human societies. New York : Norton, [2005] ©2005.

Dumor, K., Yao, L., 2019. Estimating China's Trade with Its Partner Countries within the Belt and Road Initiative Using Neural Network Analysis. Sustainability 11, 1449.

EBRD, 2020. Belt and Road Initiative (BRI) [WWW Document]. URL //www.ebrd.com/what-we-do/belt-and-road/overview.html (accessed 7.6.20).

Flint, C., 2001. The geopolitics of laughter and forgetting: A world-systems interpretation of the post-modern geopolitical condition. Geopolitics 6, 1–16.

Font Pair, 2020. FontPair helps you pair Google Fonts together. [WWW Document]. FontPair. URL https://fontpair.co (accessed 7.6.20).

Foroohar, R., 2020. Europe and US can still compete with Chinese tech [WWW Document]. The Financial Times. URL https://www.ft.com/content/cdd9322d-a0af-4bd8-b6ab-04d9fadcc301 (accessed 7.28.20).

Foster, J., 2005. From simplistic to complex systems in economics. Cambridge Journal of Economics 29, 873–892.

Frankopan, P., 2019. The new silk roads: The present and future of the world. Knopf.

Frankopan, P., 2015. The silk roads: A new history of the world. Bloomsbury Publishing.

Gabusi, G., 2017. "Crossing the river by feeling the gold": The Asian Infrastructure Investment Bank and the financial support to the Belt and Road Initiative. China & World Economy 25, 23–45.

Giles, R.H., 1970. The Role of Trade in International Relations. International Relations 3, 565–577.

Glosny, M.A., 2010. China and the BRICs: A real (but limited) partnership in a unipolar world. Polity 42, 100–129.

Griggs, T., 2020. Have the US and China passed the point of no return? The Financial Times.

Gurevich, T., Herman, P., 2018. The dynamic gravity dataset: 1948–2016. US International Trade Commission Office of Economics Working Paper.

Gurevich, T., Herman, P., Abbyad, N., Demirkaya, M., Drenski, A., Horowitz, J., Kenneally, G., 2018. The Dynamic Gravity Dataset: Technical Documentation.

Harkavy, R.E., 2013. Great power competition for overseas bases: The geopolitics of access diplomacy. Elsevier.

Harshaw, T., 2018. Emperor Xi's China Is Done Biding Its Time [WWW Document]. Belfer Center for Science and International Affairs. URL https://www.belfercenter.org/publication/emperor-xis-china-done-biding-its-time (accessed 7.14.20).

Hausmann, R., Hidalgo, C., Stock, D., Yildirim, M.A., 2019. Implied comparative advantage.

Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. Journal of international Economics 81, 1–14.

Herrero, A.G., Xu, J., 2017. China's Belt and Road initiative: can Europe expect trade gains? China & World Economy 25, 84–99.

Hess, W., Persson, M., Rubenbauer, S., Gertheiss, J., 2013. Using Lasso-Type Penalties to Model Time-Varying Covariate Effects in Panel Data Regressions–A Novel Approach Illustrated by the'Death of Distance'In International Trade.

Hille, K., Shepherd, C., Manson, K., Foy, H., 2020. US urged to exploit cracks in Russia-China relationship [WWW Document]. The Financial Times. URL https://www.ft.com/content/b59bd581-a9f8-4415-9be6-4dff722e87a9 (accessed 7.28.20).

Hodgson, G.M., 1987. Economics and systems theory. Journal of Economic Studies.

Huang, R., Nie, T., Zhu, Y., Du, S., 2019. Forecasting trade potential between china and the five central asian countries: under the background of belt and road initiative. Computational Economics 1–15.

Huang, Y., 2016. Understanding China's Belt & Road initiative: motivation, framework and assessment. China Economic Review 40, 314–321.

Hughes, A.C., Lechner, A.M., Chitov, A., Horstmann, A., Hinsley, A., Tritto, A., Chariton, A., Li, B.V., Ganapin, D., Simonov, E., Morton, K., Toktomushev, K., Foggin, M., Tan-Mullins, M., Orr, M.C., Griffiths, R., Nash, R., Perkin, S., Glémet, R., Kim, M., Yu, D.W., 2020. Horizon Scan of the Belt and Road Initiative. Trends in Ecology & Evolution. https://doi.org/10.1016/j.tree.2020.02.005

Husein, S., Ameen Kader, S., Iyer, A., Menon, A., El-Tigi, W., Rautmann, M., 2019. BRI connect: an initiative in numbers 2nd edition: Asia in focus. refinitiv.

IIGF, I.I. of G.F., 2020. Countries of the Belt and Road Initiative (BRI) – Green Belt and Road Initiative Center. URL https://green-bri.org/countries-of-the-belt-and-road-initiative-bri (accessed 7.29.20).

IMO, (International Maritime Organisation), 2004. Maritime Safety Committee - 79th session: 1-10 December 2004 [WWW Document]. imo.org. URL http://www.imo.org/blast/mainframe.asp?topic_id=110&doc_id=3665#ais (accessed 8.11.20).

Isgut, A., Ravishanker, G., Rosenblat, T., 2005. The basics of international trade: A classroom experiment. Unpublished paper, Wesleyan University.

Jacobs, L.M., Van Rossem, R., 2014. The BRIC Phantom: A comparative analysis of the BRICs as a category of rising powers. Journal of Policy Modeling 36, S47–S66.

Jervis, R., 2010. Thinking Systemically about Geopolitics. Geopolitics 15, 165–171.

Jiang, Y., Sheu, J.-B., Peng, Z., Yu, B., 2018. Hinterland patterns of China Railway (CR) express in China under the Belt and Road Initiative: A preliminary analysis. Transportation Research Part E: Logistics and Transportation Review 119, 189–201. https://doi.org/10.1016/j.tre.2018.10.002

Kaul, I., Grunberg, I., Stern, M.A., 1999. Defining global public goods. Global public goods: international cooperation in the 21st century 2–19.

Kirby, W., Van der Wende, M., 2018. The New Silk Road: implications for higher education in China and the West? Cambridge Journal of Regions, Economy and Society 12, 127–144. https://doi.org/10.1093/cjres/rsy034

Kohl, T., 2019. The Belt and Road Initiative's effect on supply-chain trade: evidence from structural gravity equations. Cambridge Journal of Regions, Economy and Society 12, 77–104. https://doi.org/10.1093/cjres/rsy036

Konara, P., Mohr, A., 2019. Why we should stop using the Kogut and Singh Index. Management International Review 59, 335–354.

Kornai, J., 1971. Economic systems theory and general equilibrium theory. Acta Oeconomica 297–317.

Kourtellos, A., Stengos, T., Tan, C.M., 2010. Do institutions rule? The role of heterogeneity in the institutions vs. geography debate. Economics Bulletin 30, 1710–1719.

Lall, S.V., Lebrand, M.S.M., 2019. Who Wins, Who Loses? Understanding the Spatially Differentiated Effects of the Belt and Road Initiative. The World Bank.

Lee, H.L., Shen, Z.-J. (Max), 2020. Supply Chain and Logistics Innovations with the Belt and Road Initiative. Journal of Management Science and Engineering. https://doi.org/10.1016/j.jmse.2020.05.001

Lee, J.-W., Swagel, P., 1997. Trade barriers and trade flows across countries and industries. Review of Economics and Statistics 79, 372–382.

Lee, P.T.-W., Hu, Z.-H., Lee, S.-J., Choi, K.-S., Shin, S.-H., 2018. Research trends and agenda on the Belt and Road (B&R) initiative with a focus on maritime transport. Maritime Policy & Management 45, 282–300.

Li, Y., Peng, Y., Luo, J., Cheng, Y., Veglianti, E., 2019. Spatial-temporal variation characteristics and evolution of the global industrial robot trade: A complex network analysis. PloS one 14.

Lippe, M., Bithell, M., Gotts, N., Natalini, D., Barbrook-Johnson, P., Giupponi, C., Hallier, M., Hofstede, G.J., Le Page, C., Matthews, R.B., Schlüter, M., Smith, P., Teglio, A., Thellmann, K., 2019. Using agent-based modelling to simulate social-ecological systems across scales. GeoInformatica 23, 269–298. https://doi.org/10.1007/s10707-018-00337-8

Liu, W., Li, X., Liu, H., Tang, Z., Guan, D., 2015. Estimating inter-regional trade flows in China: A sector-specific statistical model. Journal of Geographical Sciences 25, 1247–1263.

Lukin, E.V., 2019. Sectoral and Territorial Specifics of Value-Added Chains in Russia: the Input-Output Approach. Ekonomicheskie i Sotsialnye Peremeny 129-149A.

Magerman, G., Studnicka, Z., Van Hove, J., 2016. Distance and border effects in international trade: A comparison of estimation methods. Economics: The Open-Access, Open-Assessment E-Journal 10, 1–31.
Martin, L.L., 2015. The Oxford handbook of the political economy of international trade. Oxford University Press.
Masad, D., 2014. Simulating International Energy Security. Presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, pp. 301–308.
Mayer, T., Zignago, S., 2006. Notes on CEPII's distances measures.
Melecky, M., Roberts, M., Sharma, S., 2019. The wider economic benefits of transport corridors: a policy framework and illustrative application to the China-Pakistan Economic Corridor. Cambridge Journal of Regions, Economy and Society 12, 17–44. https://doi.org/10.1093/cjres/rsy033
Meunier, S., Nicolaïdis, K., 2005. The European Union as a trade power. International relations and the European Union 12, 247–269.
Mi, Z., Meng, J., Zheng, H., Shan, Y., Wei, Y.-M., Guan, D., 2018. A multi-regional input-output table mapping China's economic outputs and interdependencies in 2012. Scientific data 5, 180155.
Mobley, T., 2019. The Belt and Road Initiative. Strategic Studies Quarterly 13, 52–72.
Molnár, V.K., Huang, D., Hossain, M.E., 2019. Europe and Central Asia in China's Belt and Road'Initiative: A Multidimensional Analysis of Trade Intensiveness. Applied Economics and Finance 7, 33–41.
Natural Earth, 2020a. 50m-cultural-vectors [WWW Document]. Natural Earth. URL https://www.naturalearthdata.com/downloads/50m-cultural-vectors/ (accessed 8.4.20).
Natural Earth, 2020b. 10m-cultural-vectors [WWW Document]. Natural Earth. URL https://www.naturalearthdata.com/downloads/10m-cultural-vectors/ (accessed 8.4.20).
NetworkX, 2020. betweenness_centrality — NetworkX 1.10 documentation [WWW Document]. betweenness_centrality. URL https://networkx.github.io/documentation/networkx-1.10/reference/generated/networkx.algorithms.centrality.betweenness_centrality.html (accessed 8.12.20).
Newman, M., 2018. Networks. Oxford university press.
Nye Jr, J.S., 2009. Get smart: Combining hard and soft power. Foreign affairs 160–163.
Nystuen, J.D., Dacey, M.F., 1961. A graph theory interpretation of nodal regions. Presented at the Papers of the Regional Science Association, Springer, pp. 29–42.
OECD, 2018. China's Belt and Road Initiative in the Global Trade, Investment and Finance Landscape.
Ognyanova, K., 2019. Network visualization with R. Network 1, T2.
Open Street Map, 2020. Tag info, Open Street Map, place=city [WWW Document]. TagInfo. URL https://taginfo.openstreetmap.org/tags/?key=place&value=city#combinations (accessed 8.5.20).
OpenFlights, 2020. OpenFlights: Airport and airline data [WWW Document]. OpenFlights: Airport and airline data. URL https://openflights.org/data.html (accessed 8.11.20).
Oselio, B., Kulesza, A., Hero, A., 2014. Multi-objective optimization for multi-level networks. Presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, pp. 129–136.
OSM, O.S.M., 2020. OpenSeaMap: Tracking of ships [WWW Document]. URL https://www.openseamap.org/index.php?id=schiffstracking&L=1 (accessed 8.11.20).
PwC's GWC, 2016. China's new silk route. The long and winding road.
Ricardo, D., 1891. Principles of political economy and taxation. G. Bell and sons.
Robinson, J.A., Acemoglu, D., 2012. Why nations fail: The origins of power, prosperity and poverty. Profile London.
Rodrigue, J.-P., 2012. Supply chain management, logistics changes and the concept of friction. Cities, Regions and Flows 2842–2843.
Rodrik, D., Subramanian, A., Trebbi, F., 2004. Institutions rule: the primacy of institutions over geography and integration in economic development. Journal of economic growth 9, 131–165.

Rose, A.K., 2004. Do we really know that the WTO increases trade? American economic review 94, 98–114.
Ruta, M., Dappe, M.H., Lall, S., Zhang, C., 2019. Belt and Road Economics. World Bank Publications.
Sayapova, A., 2020. Regional Input–Output Tables. Studies on Russian Economic Development 31, 31–35.
Sherouse, O., 2020. wbdata: A library to access World Bank data.
Silk Road Briefing, 2020. China's Investment In Belt & Road Initiative Countries Averaging US$12.76 Billion A Month. Silk Road Briefing. URL https://www.silkroadbriefing.com/news/2020/01/14/chinas-investment-belt-road-initiative-countries-averaging-us12-76-billion-month/ (accessed 7.6.20).
Smith, A., 1776. An inquiry into the nature and causes of the wealth of nations: Volume One. London: printed for W. Strahan; and T. Cadell, 1776.
Soner, E., 2020. Effects of Belt and Road Initiative on port selection. Journal of Sustainable Development of Transport and Logistics 5.
Stanzel, A., 2017. China's Belt and Road – new name, same doubts? [WWW Document]. European Council on Foreign Affairs. URL https://www.ecfr.eu/article/commentary_chinas_belt_and_road_new_name_same_doubts (accessed 6.28.20).
Starr, H., 1991. Joining political and geographic perspectives: Geopolitics and international relations. International Interactions 17, 1–9.
Stavridis, S., 2001. Why the'militarising'of the European Union is Strengthening the Concept of a Civilian Power Europe.
Sterling, D.P., 2018. A new era in cultural diplomacy: Promoting the image of China's "Belt and Road" initiative in Asia. Open Journal of Social Sciences 6, 102–116.
Sun, Z., Yu, S., Guo, H., Wang, C., Zhang, Z., Xu, R., 2020. Assessing 40 years of spatial dynamics and patterns in megacities along the Belt and Road region using satellite imagery. International Journal of Digital Earth 1–17.
Tombe, T., Zhu, X., 2019. Trade, migration, and productivity: A quantitative analysis of china. American Economic Review 109, 1843–72.
UN Comtrade, 2020. United Nations commodity trade statistics database. URL: http://comtrade.un.org.
Valentinov, V., Chatalova, L., 2014. Transaction costs, social costs and open systems: Some common threads. Systems Research and Behavioral Science 31, 316–326.
Van Bergeijk, P.A., Brakman, S., 2010. The gravity model in international trade: Advances and applications. Cambridge University Press.
Vidmer, A., Zeng, A., Medo, M., Zhang, Y.-C., 2015. Prediction in complex systems: The case of the international trade network. Physica A: Statistical Mechanics and its Applications 436, 188–199. https://doi.org/10.1016/j.physa.2015.05.057
Wagner, S.M., Neshat, N., 2010. Assessing the vulnerability of supply chains using graph theory. International Journal of Production Economics 126, 121–129. https://doi.org/10.1016/j.ijpe.2009.10.007
Wang, C., Lim, M.K., Zhang, X., Zhao, L., Lee, P.T.-W., 2020. Railway and road infrastructure in the Belt and Road Initiative countries: Estimating the impact of transport infrastructure on economic growth. Transportation Research Part A: Policy and Practice 134, 288–307. https://doi.org/10.1016/j.tra.2020.02.009
Wang, R., Zheng, X., Wang, H., Shan, Y., 2019. Emission drivers of cities at different industrialization phases in China. Journal of Environmental Management 250, 109494. https://doi.org/10.1016/j.jenvman.2019.109494
Wang, X., 2020. Winning American Hearts and Minds: China's Image Building Efforts in the 21st Century. Springer Nature.
WB, T.W.B., 2020a. Worldwide Governance Indicators [WWW Document]. Worldwide Governance Indicators. URL www.govindicators.org (accessed 8.11.20).

WB, T.W.B., 2020b. World Development Indicators [WWW Document]. World Development Indicators. URL https://datacatalog.worldbank.org/dataset/world-development-indicators (accessed 8.11.20).

WB, T.W.B., 2020c. Logistics Performance Index [WWW Document]. Logistics Performance Index. URL https://datacatalog.worldbank.org/dataset/logistics-performance-index (accessed 8.11.20).

Wen, X., Ma, H.-L., Choi, T.-M., Sheu, J.-B., 2019. Impacts of the Belt and Road Initiative on the China-Europe trading route selections. Transportation Research Part E: Logistics and Transportation Review 122, 581–604. https://doi.org/10.1016/j.tre.2019.01.006

Wilson, A.G., 2016. Approaches to geo-mathematical modelling: New tools for complexity science. John Wiley & Sons.

Wu, D., Liu, J., 2016. Multi-regional input-output (MRIO) study of the provincial ecological footprints and domestic embodied footprints traded among China's 30 provinces. Sustainability 8, 1345.

Wu, T., Zhang, B., Cao, Y., Sun, P., 2020. Impact of Multi-Dimensional and Dynamic Distance on China's Exports of Wooden Forest Products to Countries along the "Belt and Road." Sustainability 12, 3339.

Wu, Z., Hou, G., Xin, B., 2020. Has the Belt and Road Initiative Brought New Opportunities to Countries Along the Routes to Participate in Global Value Chains? SAGE Open 10, 2158244020902088. https://doi.org/10.1177/2158244020902088

Yang, D., Jiang, L., Ng, A.K., 2018a. One Belt one Road, but several routes: A case study of new emerging trade corridors connecting the Far East to Europe. Transportation Research Part A: Policy and Practice 117, 190–204.

Yang, D., Pan, K., Wang, S., 2018b. On service network improvement for shipping lines under the one belt one road initiative of China. Transportation Research Part E: Logistics and Transportation Review 117, 82–95. https://doi.org/10.1016/j.tre.2017.07.003

Yotov, Y.V., 2012. A simple solution to the distance puzzle in international trade. Economics Letters 117, 794–798.

Yu, L., Zhao, D., Niu, H., Lu, F., 2020. Does the belt and road initiative expand China's export potential to countries along the belt and road? China Economic Review 60, 101419.

Yuan, H., 2020. The Belt and Road Initiative and international relations theories: Challenges and a new research agenda.

Zeng, J., 2019. Narrating China's belt and road initiative. Global Policy 10, 207–216.

Zeng, J., 2017. Does Europe matter? The role of Europe in Chinese narratives of 'One belt one road'and 'New type of great power relations.' JCMS: Journal of Common Market Studies 55, 1162–1176.

Zheng, H., Zhang, Z., Wei, W., Song, M., Dietzenbacher, E., Wang, X., Meng, J., Shan, Y., Ou, J., Guan, D., 2020. Regional determinants of China's consumption-based emissions in the economic transition. Environmental Research Letters 15, 074001.

8 | Appendix

A1 | Research diary

This section outlines the process of carrying out this research. The primary points of reflection revolve around meetings with Dr Thomas P. Oléron-Evans in his capacity as supervisor for this work. In addition, any other dates of significance will be added and marked.

Date	Event type	Description
07.04.2020	**Supervisor meeting 1**	Helpful kick off meeting as a group, outlining expectations, priorities and timelines. A useful session after which I could use time on hand to construct a plan and assess ethics approval for data.
16.04.2020	**Supervisor meeting 2**	Discussion of ethics and plan. All in order, little progress made due to other course commitments, the process of changing jobs. However, all issues were discussed openly and were deemed to no pose a big challenge. Next agreed milestone is to begin reading as extensively as other commitments allow and catalogue ideas and findings.
27.04.2020	**Personal reflection**	A process developed for how to effectively catalogue reading in the citation manager (Zotero) and keep track of what has been read and the key messages.
07.05.2020	**Supervisor meeting 3**	Discussion on literature encountered and the potential research agendas…received good steering in the direction of graph theory and spatial interaction modelling. Next steps continue to be on building up a view of the literature and refining the problem to address with a view build out literature review.
04.06.2020	**Supervisor meeting 4**	Based on initial literature findings that provided an overview and context for the problem and the subject of the BRI, discussing potential methods and how to solve for missing data. As the number of questions and discussion points was too large, another meeting was scheduled the following week. The aim continues to be to narrow down a method with which to address a question the data lends itself to.
11.06.2020	**Supervisor meeting 4.5**	Very useful meeting, the process of selecting a methodology and a truly specific research question has proven to be challenging. As a result, approach for the next following weeks is to focus on developing a code base and familiarity with the likely method of choice network and graph theory by referencing to CASA material from course-mates and online resources.
25.06.2020	**Supervisor meeting 5**	Time coding proved fruitful and constructed a robust base which was steadily expanded to encompass the array of notebooks on GitHub. Walk through code. Next objective is to focus on writing after focus on code created fresh headspace.
28.06.2020	**Personal reflection**	A more difficult point in time from a motivations point of view, however it was channelled into the somewhat productive endeavour of designing a front cover.
16.07.2020	**Supervisor meeting 6**	Literature review and methodology draft discussed. At this stage approximately 1 month behind originally discussed timeline but agreeing that good progress is being made.
06.08.2020	**Supervisor meeting 7**	Extension of literature review and methodology with additional code progress. Openly discussed that the subsequent 18 days will be demanding but possible.
13.08.2020	**Supervisor meeting 7.5**	Reflection on submitted draft with more detailed methodology and beginning of analysis. Very useful course correction through feedback that emphasised focus and focus of content to specific problems / research questions already identified.
21.08.2020 – 23.08.2020	**Personal reflection**	Proof reading and final edits to text and code. A harder experience than it should have been partly due to time management constraints.
24.08.2020	**Submission deadline**	Final GitHub uploads checked Work submitted for marking on UCL's Moodle page

A2 | Weighted great circle distance calculations (Gurevich et al., 2018)

Taken as is from Gurevich et al. (2018) technical documentation paper that accompanied their Dynamic Gravity Dataset. See pages 12–13 in the original document.

4.1.3 Variable Construction

The variables *latitude* and *longitude* are based on the respective locations of the largest city or cities in each country, as described in section 4.1.1. The reported values in the Dynamic Gravity dataset are the simple average of the city-level coordinates. As a result, coordinates reported for each country represent an average location for each country.

The variable *distance* is the population-weighted distance between *country_o* and *country_d*. This variable was calculated by weighting the distance between major cities of each country by each city's population. The following formula, adapted from Mayer and Zignago [2011], was used to calculate the *distance*:

$$distance_{o,d} = \left(\frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} population_i \times population_j \times distance_{ij}^{\theta}}{population_o \times population_d} \right)^{\frac{1}{\theta}} \quad \forall o,\ d \in \{countries\}$$

where i is a city in *country_o* and j is a city in country *country_d*. Note that $population_o$ and $population_d$ reflect only the total population residing in the cities of *country_o* and *country_d* that were used in calculation of *distance*, not the total population of the countries overall. We set θ, which is the sensitivity of trade flows to bilateral distance, equal to 1, which is the sensitivity used by Mayer and Zignago [2011].[14]

The distance between cities is calculated using a greater circle distance formula. Greater circle distance, or geodesic distance, is the shortest distance between two points on a sphere.[15] The formula uses the Spherical Law of Cosines to determine distance, which assumes the Earth is a perfect sphere, with a constant radius of 6,372.795 km [Weisstein, 2017]. One potential downfall of this method is that it takes the inverse of a cosine, which is ill-conditioned and can lose accuracy if the distance is small. The distances between cities are sufficiently large to avoid this issue [Chamberlain, 1996]. The benefits are that the Law of Cosines has the advantage of being computationally efficient and clear. The greater circle distance formula is:

$$distance_{ij} = K \arccos[\cos\delta_i \cos\delta_j \cos(\lambda_i - \lambda_j) + \sin\delta_i \sin\delta_j]$$

where δ is the latitude, λ is the longitude, and K is the earth's radius.

Internal Country Observations: In line with gravity trade theory, if there is only one city level observation in our data, we have assumed that internal distance is one so that the natural log of distance is well defined and equal to zero for internal trade. There are 80 countries with an internal distance of one, primarily islands or small countries (e.g. Andorra, Liechtenstein, Qatar).

A3 | Tools used

This section highlights the main tools utilised to see the project through from start to finish.

Communication with supervisor: email and Microsoft Teams

Write up: Microsoft Word as part of Office 365

Coding languages: Python

Coding environments: Majority of code was written and remains in Jupyter Notebooks. Some functions, particularly those that solve problems that needed to be addressed across Notebooks were transferred to .py packages using Atom.io and imported into the Notebook to increase readability of the process.

Visualisation: Summary charts and network diagrams are prepared in Python in their respective analysis code blocks. By exception, where multiple shape file layers were present QGIS was used. For process diagrams, draw.io was used.

Storage and hosting: code, data and the write-up itself is hosted on GitHub, throughout the development of the analysis and data extraction, updates were pushed for the sake of transparency and security in the event of unexpected loss of information locally.

GitHub repository link: https://github.com/antoniosfiala/UCL_Dissertation

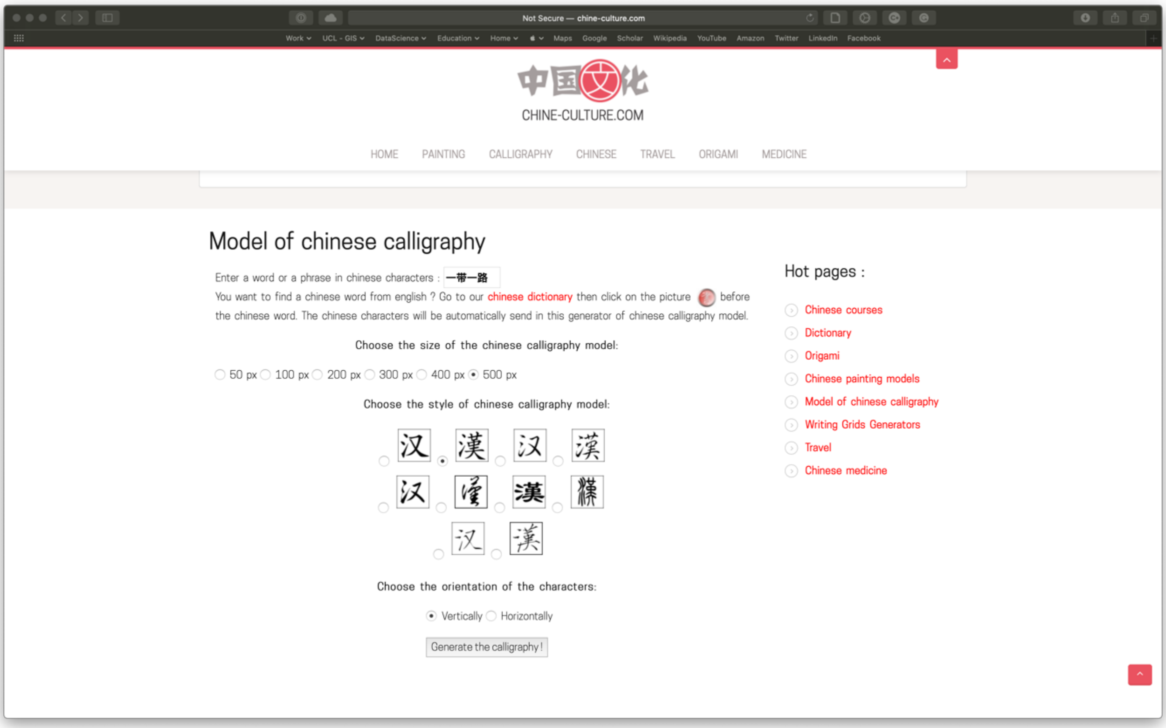
A4 | Front cover documentation

The front cover Chinse calligraphy represents the phrase 'Yidai Yilu' or 一带 一路 which literally translates into 'One Belt, One Road' and was the original phraseology in both English and Chinese for the project China unveiled in 2013. Despite the rebrand in English, the Chinese phrase remains as it was before.

The image for the calligraphy was sourced from:
http://www.chine-culture.com/en/chinese-calligraphy/model-of-chinese-calligraphy.php

The settings used were the second style from the top left at the highest resolution (500px) at a vertical setting.

A3 Figure | Illustration of the website where calligraphy was sourced (with settings)



A5 | Font selection

Font selection based on the rule of thumb for having a sans serif font as a header and a serif font for the body. To identify correct pairs, the FontPair (2020) website tool was used to make the selection for the fonts of:

- Montserrat for the headings
- Cardo for the body

Other styles considered:
- Cabin & Old Standard TT
- Open Sans & Gentium Book Basic