Article:

Tischer, D. orcid.org/0000-0003-3180-7862 (2022) Collecting network data from documents to reach non-participatory populations. Social Networks, 69. pp. 113-122. ISSN 0378-8733

https://doi.org/10.1016/j.socnet.2020.09.004

Collecting network data from documents to reach non-participatory populations

Daniel Tischer, University of Bristol (daniel.tischer@bristol.ac.uk)

Keywords: Data Collection; Social Network Analysis; Documents; Tie Strength; Careers; Biographies; Ethics;

Abstract

Collecting social network data is challenging, not least because conventional approaches rely on human participation. However, there are instances where access to research subjects is restricted or non-existent, especially in the high-stakes commercial world. This paper outlines the collection of network data from a relatively obscure financial document – the offering circular. I consider the implications of dealing with a non-participatory population and data that is not produced for social network research. In exploring the process of translating data into social network data I highlight the importance of retaining context and qualitative descriptions in the data. I also consider how different coding strategies impact the network. Finally, I discuss the triangulation, data anonymity and potential ethical-legal implications of collecting data from documents.

Abbreviations:

CDO – Collateralised Debt Obligation

CM – Collateral Manager working for CMFs

CMF – Collateral Management Firm

GFC – Global Financial Crisis

OC – Offering Circular

Introduction

Social networks data is generally collected directly from participants using interviews, observation and questionnaires; designed and deployed to answer a particular research question. This means that we often overlook other sources which contain data that can be usefully analysed using social network techniques.

This paper outlines an alternative approach to collecting social network data using documents as primary source of data. It discusses and reflects on the collection of career data for a pilot project. The data source is a specific document—the *offering circular* (OC)—published for each collateralised debt obligation (CDO), the complex financial derivative that caused the 2008 Global Financial Crisis (GFC). The OC contains detailed descriptions of the product but it also contains two sets of data that lends itself to be analysed using social network techniques:

the firms involved in the structuration of the CDO, and detailed information on the career histories of a particular financial actor called 'collateral manager'.

Collecting data primarily from documents may mitigate access constraints and resources limitations that are often burdensome obstacles in fieldwork settings, but equally, poses a number of challenges relating to data quality and the coding of data. The remainder of the document will cover these issues in turn. A first section introduces the research project. Section two explains why documents, rather than human subjects, are used to collect data. This is followed by the discussion of how we begin the data collection process. A fourth section explores preliminary decisions on data coding and how different approaches affect the network dataset. Section five discusses data triangulation as a means to ensure data accuracy and completeness. Before concluding, data annonymisation and ethical-legal concerns are considered for the use of LinkedIn data.

Project background

The data collection strategy discussed in this paper is part of a larger research project which examines actors involved in the creation of *Collateralised Debt Obligations* (CDOs), the exotic financial credit derivative that caused the 2008 (GFC). Within this broader project Author and I (Author A) complement existing literature of interconnectedness of financial market activity by adding a social lens to a field that has been dominated by economics and finance to date (see Caccioli et al. 2018; Gai & Kapadia 2010; Haldane & May 2011). We do this by tracing the complex social relations that developed between financial organisations involved in the creation of CDOs (Author B).

The pilot project discussed here zooms in one specific actor: 'collateral managers'. To avoid confusion, the term *collateral management firm* (CMF) is used to refer to the firm, whilst the term *collateral manager* (CM) is used in reference to individuals employed by these firms. CMFs are interesting financial actors because they sit between investors (buyers) and investment banks (sellers) and have a key risk-mitigating function: they are hired to select and manage assets on behalf of the investor (Author A). But in the aftermath of the GFC it transpired that this was not always the case. Numerous legal actions pursued by the SEC reveal that several key CMFs have neglected their fiduciary duties by making decisions that benefit the sellers rather than the investors (see SEC 2016 for an overview; c.f. Bavoso 2017). Most legal cases place emphasis on the relationships and interactions between IBs and CMFs which created conflicts of interest as it was investment banks who hired the CMFs. Nonetheless, CMFs claimed to act independently from investment banks for the benefit of investors (Author B and Chernenko 2017). For example, one CMF claimed to have "no traditional ties to investment banking" (Jupiter Highgrade II CDO, 2004a: 25), despite employees previously working for several investment banks. In another case, Goldman Sachs contracted CMFs that were founded by and staffed with former Goldman Sachs employees: *Greywolf Capital Management* and *GSC Partners.* In both cases Goldman Sachs shorted over a third of assets, that is, Goldman bet on the default of these assets to profit with investors incurring high losses on their investments (US PSI 2011: 392f.).

The aim of the research is thus to trace the entanglement of this particular, supposedly independent financial market actor with Wall Street investment banks by studying the pre-existing social ties derived from CM career biographies; and, to explore how this may have contributed to failure of the CDO market and the ensuing GFC. Rather than focusing on board interconnections, alliances or resource flows, we use the individual careers to gain a broader understanding on how knowledge, experience and personal relationships *connect* collateral management as activity and financial firms. Studying careers in the context of social network research is, of course, not ground-breaking given Granovetter's (1974) seminal work *Getting a Job* and the many subsequent papers examining "careers" (of which there are 146 in *Social Networks* alone). But in this project the treatment of careers transcends the individual and explores the entanglement of CMFs with Wall Street and the finance industry. That is, the project does not focus on the CM career as such, but rather focuses on how these individuals connect a large number of financial firms.

Why do I collect data from documents and not the actors?

Commonly used and widely discussed approaches to collecting social network data in fieldwork settings—interviews, observation, or questionnaires—(c.f. Borgatti et al. 2013; Crossley et al. 2015; Scott 1991; Wasserman & Faust 1994) may not be suitable in all research contexts, including this study. Participatory data collection requires participants, and (financial) firms in general are notoriously difficult to recruit for research projects (Author A, Roden et al 2020). Even where set boundaries limit the scope and scale of the undertaking, it may still be challenging to 'reach' employees because of geography, hierarchies, and commercial sensitivity (Borgatti et al 2013; Useem 1995). This is because access needs to be negotiated at various levels – individual, team and organisation – and often requires significant commitment and support from managers. Where network scholars manage to negotiate access to multiple organisations, coordination and resource limitations remain an issue.

Recruiting collateral managers is even more difficult given that 1) they are relatively small, 'boutique', actors and interested in retaining their privacy; 2) the research setting is not contemporary but historic; and, 3) CMFs and CMs have been subject to intense legal and regulatory scrutiny which may naturally limit their interest in collaborating with external researchers (Crossley et al. 2012, van der Hulst 2009). Taken together, these factors restrict our ability to research these actors even if we were able to recruit "a seed set" of actors (see Stys et al. 2020 for further commentary on difficult to reach populations). The fact that this research is historic—it examines the period leading up the GFC of 2007/08—further complicates collecting data in person not least because many of the firms are defunct or people have moved on to work elsewhere.

Against backdrops like these, documents become a key ally in our quest to research non-participatory populations and are often used by historic network researchers (see, for example, Breure & Heiberger 2019 or Elo 2018). The fact that documents are now available in a digital format, and often exclusively so, provide us with a readily accessible "point of entry into contemporary problems" (Riles 2006: 2) and repositions documents as more than simply a source of content, but as quintessentially involved in organising activity (Preda 2002; Prior

2008). Contracts serve as prime example of 'creating' and 'documenting' relations: they contain detail on a particular transaction between two (or more) parties and structure the relationship between those parties or may set out procurement terms between actors, public and private (Reeves-Latour & Morselli 2017). In other words, contracts present both technical systems and communities of discourse (Suchman 2003). Documents are fixed entities; or better said, their content becomes fixed when published and this durability, or immutability, of the document is what makes it a valuable source of (historic) data (Shankar, Hakken & Osterlund 2017). As Rowlinson et al. (2014) argue, (organisational) documents are not only drivers of (organisational) narratives, but they are important, albeit "under-utilized" artefacts for constructing data (Riles 2006).

Offering circulars (OCs) present such a key documentary source of data. Not only do they set out the 'terms of trade' and offers prospective buyers a detailed description of the product, payments schedules, conflict resolution, etc.; they also provide background information for the CMFs' history and ownership as well as detailed career biographies of the employees, CMs, including information on educational attainment, previous employments and specific expertise gained (Author A). This data is only available for CMs, but not for the investment bankers, or *any* other actors involved in the structuration of CDOs. This places emphasis on the CMFs' central role in 'getting the deal done', not least by advertising their expertise to potential, preselected investors, and credit rating agencies (Author A; Chernenko 2017).

It is important to point out that the intended audience for OCs is very narrow. They serve as a reference point for parties to the trade but were not written for wider public consumption. In that sense they are unlike newspaper articles, annual accounts or parliamentary proceedings. OCs do not encourage action, but silence the reader into accepting rather than questioning the content (Authors A). They are inward facing, laden with technical-legal jargon and rather repetitive and boring. They silence the reader rather than prompting interest of action These documents enter the public space almost by accident. They are not kept in a public database but enter the public space only because the CDO product requires 'listing' on stock exchanges to become investable in by pension funds and other investors.

Collecting network data from these documents is not that dissimilar to more widely discussed archival network research (see Marsden 2005, Borgatti et al. 2013). Yes, the absence of a formal archive suggests that OCs are not meticulously curated for archival purposes as was the case with Padgett and Ansell (1993) 'Florentine Families', but that does not make them any less valuable for data collection purposes. In fact, whilst Padgett and Ansell worked with multiple types of documents—historical books, tax assessments and neighbourhood co-residence—to identify actors and establish relations between them (Crossley et al. 2015), this study benefits from building the network from one type of document which enhances the consistent and commensurability of the network relations created.

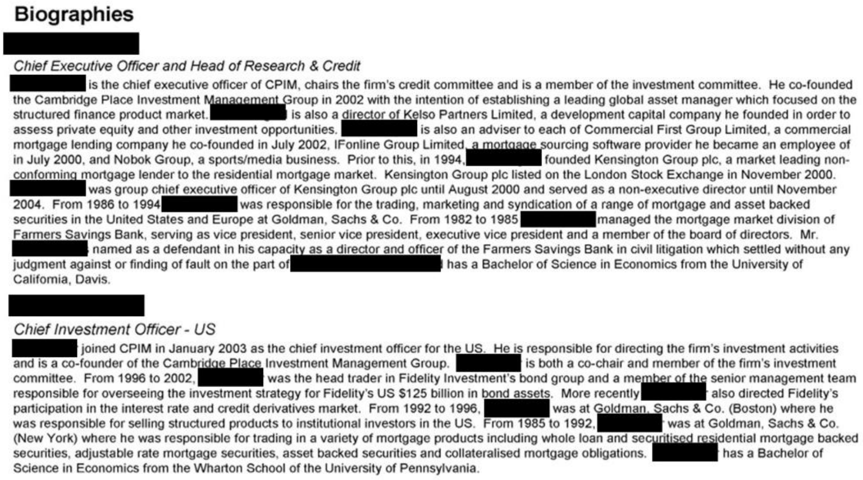
The ability to re-interpret and re-associate data contained in the OC is important to this study (Shankar et al. 2017). Offering circulars where initially published as reference of a CDO trade between various parties. But since then, the GFC has changed how we, the public, think about CDOs and documentation—OCs, pitchbooks and term sheets—have been referenced widely in public investigations and litigation of investment banks, collateral managers and hedge funds (see FCIC 2011 and SEC 2016). The GFC elevated these documents from depictions of mundane technical-legal practices to records of a financial crisis in the making. In other words, these documents became artefacts of the GFC and the re-examination of their content is not only warranted but necessary (Scott 1990: 34f.).

Collecting network data from documents

An initial point to be made is that the researcher is not participating in the creation of the data but is presented with an existing and 'finite' set of relations and attributes. While it may be possible to collect additional data from other sources, the data collection starts with the document, the offering circular, and the data it contains. Unlike primary data collected by researchers with a specific analysis in mind, OCs and most other documents are not produced for academic research. This means that researchers have "minimal [or no] influence in the production of data", however, decisions made by the researcher when extracting data—what to include and what to exclude—will impact the research (Crossley et al 2015: 66). It also means that the researcher needs to familiarise themselves with the data as presented, *post hoc*, rather than developing and testing interview guides and measurement scales.

A first step, therefore, requires the researcher to develop an understanding of the data contained in these documents. This pilot study of collateral management focused on 25 CMFs who listed 592 individuals as CMs. These CMs worked for 717 different firms prior to working at a CMF. Upon first inspection OCs appear are relatively standardised: 1) in terms of the fundamental document architecture; 2) the technical-legal jargon they are written in; and, 3) the detail they provide (Author). Biographies for each individual feature a short paragraph detailing five basic data categories: name, university level education, previous employment including roles held, and the CMF they are working for (Figure 1). These data points feature consistently across OCs and the CM biographies contained therein. As illustrated in Table 1, data was collected as egonet data where the CM (ego) is connected to a set of firms (alters) (Crossley et al. 2015). It is time-stamped and spans the period from education to the current employer, the collateral manager. The data is presented as a two-mode or affiliation network where each CM has a set of relationships with organisations they previously worked for, plus the collateral management firm.

======= Figure 1 about here =========



Biographies

[redacted]

Chief Executive Officer and Head of Research & Credit

[redacted] is the chief executive officer of CPIM, chairs the firm's credit committee and is a member of the investment committee. He co-founded the Cambridge Place Investment Management Group in 2002 with the intention of establishing a leading global asset manager which focused on the structured finance product market. [redacted] is also a director of Kelso Partners Limited, a development capital company he founded in order to assess private equity and other investment opportunities. [redacted] is also an adviser to each of Commercial First Group Limited, a commercial mortgage lending company he co-founded in July 2002, IFonline Group Limited, a mortgage sourcing software provider he became an employee of in July 2000, and Nobok Group, a sports/media business. Prior to this, in 1994, [redacted] founded Kensington Group plc, a market leading non-conforming mortgage lender to the residential mortgage market. Kensington Group plc listed on the London Stock Exchange in November 2000. [redacted] was group chief executive officer of Kensington Group plc until August 2000 and served as a non-executive director until November 2004. From 1986 to 1994 [redacted] was responsible for the trading, marketing and syndication of a range of mortgage and asset backed securities in the United States and Europe at Goldman, Sachs & Co. From 1982 to 1985 [redacted] managed the mortgage market division of Farmers Savings Bank, serving as vice president, senior vice president, executive vice president and a member of the board of directors. Mr. [redacted] named as a defendant in his capacity as a director and officer of the Farmers Savings Bank in civil litigation which settled without any judgment against or finding of fault on the part of [redacted] has a Bachelor of Science in Economics from the University of California, Davis.

[redacted]

Chief Investment Officer - US

[redacted] joined CPIM in January 2003 as the chief investment officer for the US. He is responsible for directing the firm's investment activities and is a co-founder of the Cambridge Place Investment Management Group. [redacted] is both a co-chair and member of the firm's investment committee. From 1996 to 2002, [redacted] was the head trader in Fidelity Investment's bond group and a member of the senior management team responsible for overseeing the investment strategy for Fidelity's US $125 billion in bond assets. More recently [redacted] also directed Fidelity's participation in the interest rate and credit derivatives market. From 1992 to 1996, [redacted] was at Goldman, Sachs & Co. (Boston) where he was responsible for selling structured products to institutional investors in the US. From 1985 to 1992, [redacted] was at Goldman, Sachs & Co. (New York) where he was responsible for trading in a variety of mortgage products including whole loan and securitised residential mortgage backed securities, adjustable rate mortgage securities, asset backed securities and collateralised mortgage obligations. [redacted] has a Bachelor of Science in Economics from the Wharton School of the University of Pennsylvania.

Figure 1: Anonymised excerpt of collateral managers' careers biography

Source: Camber 3 pitchbook (2005)

======= Table 1 about here =========

Career	Firm	Position	Duration	Ends
Individual	Prudential Securities, Inc.	Senior Research Analyst	2	1996
	Salomon Brothers	Senior Research Analyst	3	1999
	New York Life Insurance Company	Portfolio Manager	2	2001
	SG Gowen	Senior Trader	1	2002
	Nomura Securities	Senior Trader	2	2004
	Maxim Advisory LLC	Managing Director	2	2006
	Harding Advisory	CEO	2	2008
Education				
BA	Econonmics	University of Rhode Island	1988	1988
MBA	Finance	Babson College	1993	1993
CMF Information	Established	#of CDOs		
Harding Advisory	2006	22		

Table 1: Collected data for one CM career

A closer reading, however, highlights subtle differences across the description of careers summarised in Table 2. First, there is *variation in the number* of biographies included. Some OCs only contain 10 biographies whilst others feature 30 or more individual biographies. This could be a sign of over- or under reporting, however, it is also the case that CMF can be broadly split into two types: 1) a large number of CMFs have the status of "boutique" investment firms—small actors specialising in one or few financial market activities—and therefore tend to have fewer employees; and, 2) subsidiary undertakings of large established firms with broader connections across financial markets (Chernenko 2017). The number of biographies listed may reflect this distinction but it is also possible that CMFs omit information and others provide overly elaborate accounts and numbers of people to establish the collective expertise of CM employees. For example, *Aladdin Capital Management* only lists 'key personnel', whereas *Declaration Management & Research* features 24 employee profiles grouped by activity: 'structured products', 'CDO investment and trading', 'fundamental research' and ‘quantitative research’. This variation could be problematic; however, it is a problem that researchers of interlocking directorates commonly deal with by cleaning data systematically (Heemkerk et al. 2017). One way of resolving this variation may be by excluding more junior and administrative functions from the analysis as these have little input in the operational management of collateral assets. However, any such decision should only be taken at a later stage and for now, biographic data for all employees should be collected.

=======Table 2 about here =======

	minimum	maximum
# of career biographies	7	45
Completeness	limited to key senior executives	comprehensive - includes all personal across functions
Detail on Education	basic level of degree and institution	information on topic, year of graduation, institution and city
Detail on Employment	simple listing of key positions	detailed discussion of previous positions with tenure/dates
Quality of information	generic descriptions of role	scale and scope of current and previous roles

Table 2: Range and detail of data available for different collateral management firms

Second, the *detail of information* for each individual may differ across documents. As a minimum detail on education includes degree type [BA, MSc, PhD] and the institution from which it was obtained. More elaborate accounts also include the programme and year in which the degree was obtained [MA in Sociology 1987 or PhD in Nuclear Physics 2004]. Similarly, some biographies offer more detail on job descriptions than others. Biographies may include a sentence on responsibilities or achievements, but some more formulaic accounts only disclose “worked at x”. A typical format of descriptions that may be found in these documents is illustrated in Figure 2a. Here information can be directly transferred into data. However, in few cases, data ambiguity has emerged as an issue. As illustrated in Figure 2b, information may be unclear and require some form of translation by the researcher. In the example, information on job tenure is ambiguous and, in the absence of additional information, multiple options to code that data require consideration. The lack of specificity on how the “six years” are split between these two firms.

=========Figure 2a and 2b about here=======

“Prior to joining [CMF 1] in 2003, [Person A] spent 10 years at [Firm X], most recently in [Role R1]”

Name of CMF	Name of Employee (CM)	Previous Employer	Tenure of Employment	Start of Employment	End of Employment	Position/Role
CMF1	Person A	Firm X	10 years	1993	2003	Role R1

Figure 2a: A simple, one-option example of capturing data

“[p]rior to joining [CMF1] in 2004, [Person B] worked for six years at [Firm Y] and [Firm Z]"

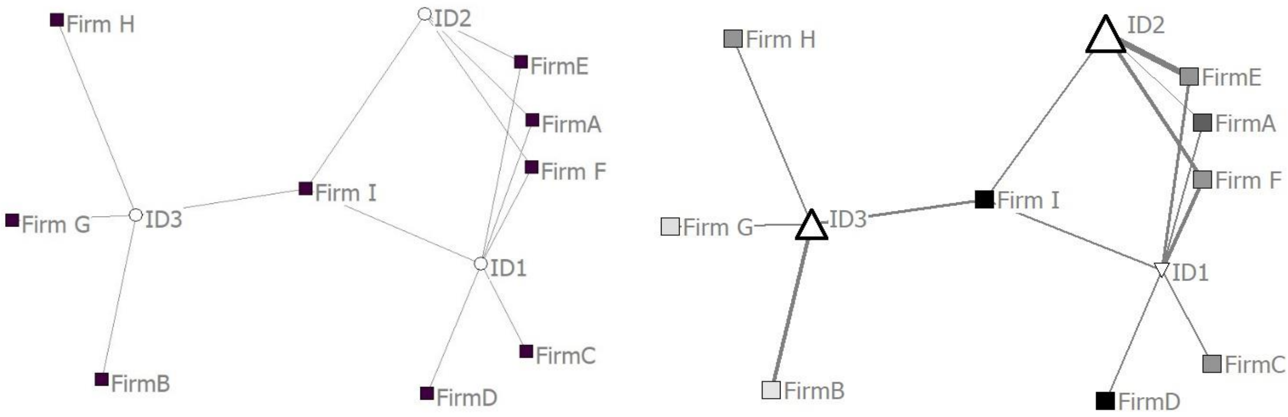
	Name of CMF	Name of Employee (CM)	Previous Employer	Tenure of Employment	Start of Employment	End of Employment
Option 1	CMF1	Person B	Firm Y	6 years	1998	2004
	CMF1	Person B	Firm Z	6 years	1992	1998
Option 2	CMF1	Person B	Firm Y	3 years	2001	2004
	CMF1	Person B	Firm Z	3 years	1998	2001
Option 3	CMF1	Person B	Firm Y	1-5 years	1999-2003	2004
	CMF1	Person B	Firm Z	1-5 years	1992-2002	1999-2003

Figure 2b: An example of options to capture ambiguous career data

It is worthwhile noting that these cases of data ambiguity are relatively rare occurrences: in the pilot study, these occurred in approximately 1-2% of biographies. In addition, Heemskerk et al (2017: 8) note that incorrect edge values may impact the analysis; however, data “volume will counter the effect of missing or incorrect data values.” Still, there is no straightforward solution to account for this ambiguity and either of the options presented provide a set of operational problems. In most of these cases, data triangulation resolves the ambiguity, however, where this is not the case, one option may be to code data in accordance with the second option and to split tenure between the two firms. This may have implications for the discussion of tie strength and needs to be taken into account when analysing the network data. However, the impact on the overall network structure is thought to be minimal given the rarity of this event occurring.

In the first instance it is important to maximise the data collected for each individual to allow for decisions about coding to be made later on. Figures 3a and 3b illustrate the difference between networks constructed from basic and comprehensive data collected. Figure 3a illustrates basic connectedness in structuralist terms. It only contains ego-alter connections with all three alters sharing ties to Firm I. A result of only connecting base data would be that any further analysis would be very limited. Figure 3b on the other hand provides a more complex story. All individuals work at Firm I, the CMF. ID3 has previously worked at two credit rating agencies (Firms B & G) before working at an investment bank and the specialist technical knowledge gives it a senior position. ID1 and ID2 both began their career in real estate prior to working in investment banking. ID1 worked for another CMF (Firm D) but left it after a short period (a negative move). ID2 moved directly from an influential investment bank (Firm E) for which they worked for a long time giving it a senior role at the CMF. The presence of additional data as represented in Figure 3b provides a much richer account of these careers which can be explored in the analysis.

=========Figure 3a and 3b about here (arrange side by side) =======



- Line thickness = tenure
- Positive (triangle up) or negative (triangle down) follow-up moves
- Firm specific characteristics (bank = grey; CM = black; CRA = light grey; real estate = dark grey)
- Seniority in years (ID node size) and role (ID node rim)

Figure 3a and 3b: Illustration of detail available for basic and comprehensive data gathering

The data collected from the career biographies included in the CDO document is not dissimilar to data on interlocking directorates. Interlock data can be obtained from *BoardEx* or *ORBIS*, which essentially outsources the process of data collection (Heemskerk et al. 2017; Valeeva, Heemskerk, Takes 2020). However, data collected manually from documents gives researchers additional contextual knowledge (Heemskerk et al. 2017: 5) which may help to alleviate some of the concerns raised by Mizruchi (1996), including the potential overestimation of the impact of director interlocks on firm behaviour and performance. The inclusion of less-senior employees working for CMFs provides an interesting account of interfirm networks that is generally ignored by the more abstract notion of board interlocks. As Godechot (2008) shows, employees of financial firms are particularly mobile because their relational assets and technical expertise are easily transferred to other financial firms who seek to benefit from them. As noted by the FCIC (2011: 139), it was important for investment banks to have “existing relationships with collateral managers”, especially once the market for CDOs showed signs of distress (SEC 2010: 7f.). Therefore, the focus of this analysis on traders and technical experts provides valuable insights into how knowledge and experience are transferred between financial firms (cf. Everett et al 2018).

There is also value in retaining the order in which individuals move between firms and integrating some form of sequence analysis as done by Bison (2014) may enhance our understanding of the interconnectedness of financial actors. For example, we may find that some people are drawn into collateral management from related (insurance) or unrelated (real estate) sectors and specialist skills or industry knowledge may constitute key reasons for hiring these individuals. Or CM employees may have acquired insider knowledge by working for credit rating agencies whose key responsibility was to rate the default risk of CDOs (FCIC 2011). Being employed by an investment bank to structure CDOs prior to being hired by a CMF may equally be a relevant consideration, especially if the CM is then hired by the investment bank. Hence, we may want to consider the second to last career step—the one prior to working for a CMF—to be of higher significance than preceding employment. Alternatively, length of tenure may be considered by weighting ties.

How do different approaches to collecting and coding data impact on the network? A preliminary view

Coding career data appears reasonably straightforward: ego working for Firm A for eight years would receive a tie strength of "8". Given that we focus on careers and ties between organisations and those who were employed there, we may want to consider the length of tenure as an indicator of the strength of ties (Granovetter 1973; 1985). At the level of ego, a 15-year tenure at a firm may be considered 'stronger' than a 1-year term. All tenures lasting >5 years may be coded as strong ties with the value '3', and short tenures of less than a year may carry the value '1'. In certain situations, ties may carry a negative weighting of '-1', for example, if we hypothesise that very short tenures have a negative influence on social capital (Borgatti & Everett, 2014; Labianca & Brass 2006). But there is danger in only considering tenure as an indicators of tie strength as other indicators, such as seniority or role, may equally (co-)determine the strength of the relationship (Table 3).

==========Table 3 about here=========

	STRONG TIES	WEAK TIES	NEGATIVE TIES
INDIVIDUAL LEVEL			
Tenure	cut-off point context-specific but possibly larger than 5 years	short employments in the firm that limit getting to know and becoming known in the firm	very short employments, e.g. shorter than 1 year may be considered opportunistic
Seniority	senior level employees may be connected to decision makers or able to take decisions themselves	juniority may limit ability to interact across the organisation	failure to be promoted may reflect negatively on the individual and its relationship with the firm
Role Embeddedness	leading teams or managing departments may lead to stronger connections with employees	limited contacts across the firm may restrict the networking opportunities for individuals	technical expertise may isolate individual from other employees
Follow-up move	move can be seen as providing future opportunities that may strengthen tie	move to unrelated function may weaken ties immediately or over time	move to direct competitor may turn a previously positive relationship negative
COLLECTIVE LEVEL			
Multiple indicators	non-compete move of multiple senior employees or teams with long tenures and high role embeddedness	random moves of individuals between firms at different stages of their careers	team moves under relatively senior person from firm to direct competitor

Table 3: Considerations of tie strength at different levels

To assess how different ways of recording data can impact on the network, five different coding strategies are considered here (see Table 3). They are used to illustrate the point that coding matters, rather than providing an analysis of the pilot data set. Both the *Tenure w/ 10y* and *w/ CMF +2* are based on the premise that more distant employments become obsolete.

=========Table 4 about here ===========

	Approach	data coded as	Network Dimensions	Loss of events / firms

Base scenario: Tenure	coding of data as number of years at a firm	valued	592 X 717	-
Binary	tie strength removed	0/1	592 X 717	-
Tie strength categorical	categorises ties by strength	"3" = x > 4; "2"= 1 > x ≤ 4; "1" = x ≤ 1	592 X 717	-
Tenure w/ 10y	only codes employee positions within the past 10 years	valued	592 X 472	245
Tenure w/ CMF +2	only codes two positions held prior to, and CMF position	valued	592 X 509	208

Table 4: Overview of multiple coding strategies

A first observation relates to the dimensions of the networks. *Tenure*, *Binary* and *Tie strength* all have 592 egos (employees) and connect to 717 events (firms). However, the network dimensions are reduced for firms by approximately 1/3 when we decide to remove older or 'latent' employments at a fixed distance—the CMF position and the two preceding employments)—or limited to a certain period, say 10 years. *Tenure w/ 10y* loses more firms as some actors have been in their recent roles for over ten years which effectively removes all prior employments form the dataset. Limiting careers that are taken into account by years or number of previous positions has the effect that network density increases slightly (0.022 compared to 0.018 for *Tenure*).

Comparing the five matrices using the quadratic assignment procedure highlights the impact on the network structure. Using *Tenure* as a base matrix we find that both matrices containing binary and categorical coding are significantly different with correlations of .71 and .86 which would be in line with our expectations. Comparing *Tenure* with *Tenure w/ 10y* and *w/ CMF +2* however shows that despite the removal of firms, the matrices are still highly correlated (.92 and .94). This is likely because many firms that are omitted are only weakly connected in the network and link to roles taken by senior employees a long time ago (some of the individuals have careers lasting over 40 years).

=========Table 5 about here ===========

	Binary	Tie strength categorical	Tenure	Tenure w/ 10y	Tenure w/ CMF +2
Binary	1.00				
Tie strength categorical	0.93	1.00			
Tenure	0.71	0.86	1.00		
Tenure w/ 10y	0.61	0.75	0.92	1.00	
Tenure w/ CMF +2	0.63	0.78	0.94	0.94	1.00

Table 5: QAP correlations for employee-firm network (all significant at .0002)

However, choices made during data collection and coding also impact the actors directly. Firm level membership in the top 10 is limited to banks and collateral management firms. *Tenure*, *Binary* and *Tie strength* accentuate banks, whereas *Tenure w/ 10y* and *w/ CMF +2* place more emphasis on CMFs, presumably because of the removal of CMs' early-career employments.

Nonetheless Top 10 membership is stable with seven actors (6 CMFs and 1 bank) present in all networks except for *Tenure w/ 10y*. The presence of CMFs in the top 10 in the pilot network is determined by the number of employers, e.g. the largest CMF lists 47 employees. However, collateral managers will become less prominent in the final study as their prominence is limited to the number of CMs working for them as named in the OCs. Banks on the other hand will become more prominent in the network as more CMFs are included. The three largest banks—Merrill Lynch, Credit Suisse First Boston and Goldman Sachs—previously employed 37, 35 and 31 CMs with a collective tenure of 164, 120 and 210 years respectively.

=========Table 6 about here ===========

Employee-firm (2mode)	Top 10 membership		Top 10 degree	
	Banks	Collateral Managers	Min	max
Binary	4	6	24	47
Tie strength categorical	4	6	52	122
Tenure	2	8	147	382
Tenure w/ 10y	1	9	137	376
Tenure w/ CM +2	1	9	137	376

Table 6: Comparison of top 10 membership across networks for employee-firm networks

Early reflections on interfirm ties

The data collection and coding undertaken in the pilot project is of particular importance to this study because the final network analysis focuses on interfirm, and not individual CMs *per se.* That is, unlike Granovetter's (1974) study, the aim is not to understand network outcomes on individual careers, but rather how numerous careers across a specific activity influence firm and market behaviour through knowledge transfers etc. We know from previous studies in economic geography (Beaverstock & Hall 2012; Beaverstock, Falcounbridge & Hall 2014) that employees in finance move between similar firms. Knowing and anticipating this degree of overlap enables the construction of whole-network structure (Perry, Pescosolido & Borgatti 2018). As individuals move between firms, they create connections, and multiple individuals moving between two firms create multiple ties between those firms. Because of this, we may need to consider not only how many ties exist between two firms; but also the strength of ties of individual employees which may influence tie strength between firms. Can two-mode whole network simply be transferred into a one mode network to exhibit relationships between firms derived from CMF employees' moves (Crossley et al 2015: 17)?

Tables 7 and 8 illustrate the impact of the decisions made about the coding and exclusion of careers step have on the final network. They exacerbate previous disconnects between the data and therefore how we think about the transformation of employee-firm ties into interfirm ties may require the final dataset to be analysed further before deciding on the right approach.

=========Table 7 and Table 8 about here ===========

	Binary	Tie strength categorical	Tenure	Tenure w/ 10y	Tenure w/ CM +2

Binary	1				
Tie strength categorical	0.89	1			
Tenure	0.62	0.83	1		
Tenure w/ 10y	0.55	0.66	0.67	1	
Tenure w/ CM +2	0.53	0.69	0.84	0.73	1

Table 7: QAP correlations for interfirm networks (all significant at .0002)

Inter-firm (1mode)	Top 10 membership		Top 10 degree	
	Banks	Collateral Managers	Min	max
Binary	7	3	70	106
Tie strength categorical	5	5	263	432
Tenure	4	6	1038	2417
Tenure w/ 10y	4	6	478	970
Tenure w/ CM +2	3	7	655	1395

Table 8: Comparison of top 10 membership across networks for inter-organisational networks

We can think of individuals in terms of 'flows' between firms. Whilst formally disconnected that bring previous experiences and knowledge to their new firm but may also retain their existing network which may not create 'direct' reciprocity but other forms of reciprocity as discussed by Baker (2014). Litigation data, as previously noted, illustrates that favour-making does influence CMF selection in some cases, but whether this is a more structural feature remains to be seen. Team moves, i.e. groups of interconnected individuals moving between two firms may be a strong indicator for this occurring. Or, as noted by Mizruchi (1996) it may allow us to zoom in on particular relationships which may evidence illicit practices, such as *collusion* or *co-optation*. More generally, focusing on interfirm ties presents opportunities to develop a more in-depth understanding of collateral managers entanglement with other financial actors, not through financial flows, but rather through personal relations.

Hence a simple transformation of data from two mode into one mode data may be insufficient. Figure 4 provides a schematic overview of how ties between firms may be constructed in the final study to take a more sophisticated approach and to account for multiple employees moving between two firms. Scenario 1 stipulates that firms connected through multiple employees who have strong ties with firms may themselves exhibit a strong tie. *Vice versa*, multiple individuals moving between two firms to which they are weakly connected may not create meaningful or strong connections between these firms.

=========Figure 4 about here ===========

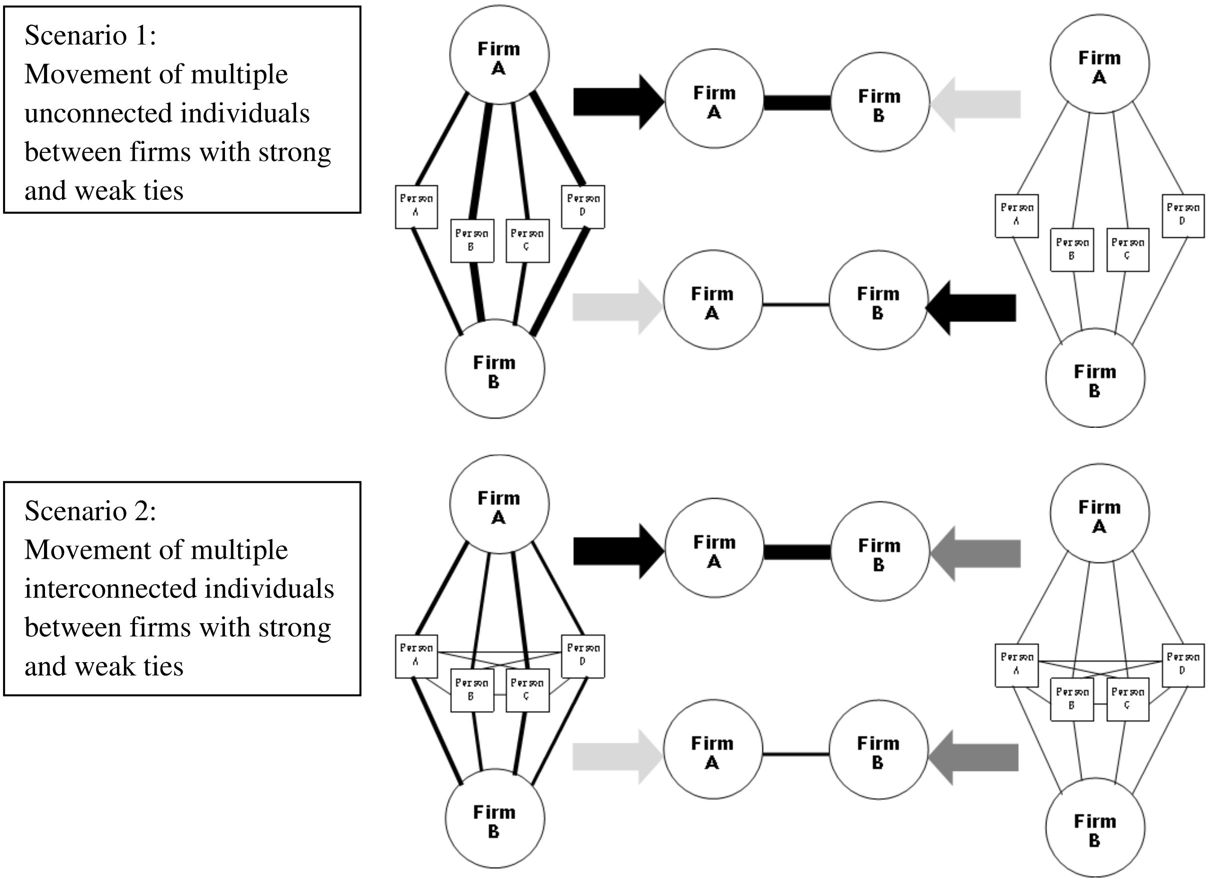


Figure 4: Strong and weak tie scenarios for multiple connected and unconnected individuals moving between firms (line thickness = tie strength; Colour of arrows signals chance of inter-firm tie strength (dark = likely; light = unlikely)

Scenario 1 assumes that individuals are not directly connected as they may have worked for firms at different times or in different functions. Scenario 2 focuses on what Godechot (2008) describes as *team moves* between firms. These connections are illustrated as lines between individuals. There are instances in the dataset where a group of employees from one firm move to or establish a collateral management firm. The prime example is a move of key Goldman Sachs personnel to *Greywolf Capital Management*: seven of the nine employees listed have worked at Goldman just before launching Greywolf. Such a scenario is qualitatively different to that described in scenario 1 and may indicate a continued working relationship between Greywolf and Goldman Sachs (Reuters 2010). Where ties between individuals are weak, it is more difficult to discern if this has positive or negative implications for tie strength, nonetheless, it does not automatically preclude the existence of strong ties between organisations.

Still transforming individual careers into interfirm ties ought to impact the analysis, not least because of the work that goes into transforming the data as collected into interfirm ties. What appear to be minor decisions made when collecting data or early on in the process of coding the data may be become amplified in the final interfirm network.

Triangulating data

Collecting data from technical-legal documents such as offering circulars should yield reliable data, not least because these documents are vetted by legal advisors. But no dataset is perfect and data collection in network research may suffer from informant errors and recall bias with a preference for more "frequent, intense, and recent contacts" (Marsden 1990: 447f.). Or data may be knowingly incomplete, as is the case in director interlocks discussed by Mizruchi (1996) and Heemkerk et al. (2017). Rather, data triangulation may be used to *cross-validate* data and to maximise data completeness (Flick 1992). The focus here is on the individual biography and the corroboration of the data, rather than locating additional data not already exhibited in the offering circulars or data for additional individuals.

Locating alternative sources of data for triangulation does of course depend on the topic of the research and may provide opportunities to gather data via interviews or questionnaires to ensure data validity over time and space (Cohen, Manion & Morrison 2007: 143). However, there is limited opportunity to cross-validate the data directly with employees as these are considered difficult to reach populations (Bailey 1994: 294ff.) and the accounts are historic. The aim of triangulation here is to ensure data collected is as complete as possible by identifying alternative sources of data: databases, archives, quality news outlets, company documents etc. Given that this research project uses data on individual careers, a key source of additional data is LinkedIn, a globally active professional network with over 650 million users. Crucially, LinkedIn user profiles contain similar data points—education, employment history, location—and thus are most suitably sources for comparison.

One key reason why LinkedIn works well in this case is that there is a set of pre-identified individuals we are seeking to validate data for. But this is only possible because we have already identified individuals. LinkedIn may be less suitable if this was not the case or as a primary source of data. For example, without knowing the names of individuals, a LinkedIn search for "collateral management" yields over 750,000 results, and many of these are irrelevant to the structuration of CDOs and of more general nature. Searching for "collateral management CDO" yields over 3000 results, but it is not easy to discern if these individuals have, in fact, worked for CDO CMFs. Additionally, these number include individuals that joined CMFs after 2008 in addition to those that took part in structuring CDO prior to the GFC in 2008. Searching for CMF directly equally yields too many or irrelevant individuals. Searching for "CVC Credit Partners" for example generates over 200 results, and the CVC LinkedIn page refers to 134 current employees[1].

Triangulating data using LinkedIn is relatively straightforward in procedural terms as we have key information for each individual CM (name, employer and education). Using these parameters, we can locate CMs and collect data manually[2]. Whilst many LinkedIn profiles corroborate the biographies provided in OCs, it is useful to discuss examples where information differs between the document and the LinkedIn profile. For example, a general degree

[1] Data accurate as of March 2020
[2] It is important to note, that LinkedIn does not allow the use of software to scrape or copy data from its website - see https://www.linkedin.com/legal/user-agreement#dos for further information

description—sociology, mathematics, statistics—may be specified as "Political Science & Government" or "Financial Risk Management", or, *vice versa* a PhD in "Control and Dynamic Systems" may become a PhD in "Mathematics". Where two descriptions are offered, it seems appropriate to collect both descriptors, before cleaning and categorising the final dataset in a unanimous way. In some cases, individuals may not be found on LinkedIn, but can be located on other services instead, such as *Bloomberg Executive* or corporate websites. The process of collecting this data is similar as for the discussion above.

Subject anonymity and ethical-legal considerations

Data collected from offering circulars is not problematic in terms of ethics as the data is public. Nonetheless we should recognise individuals' right to privacy and take steps to ensure that subjects cannot be identified even where data is publicly available as the data was not intended for these purposes. Therefore, retaining people's privacy is crucial to avoid harming these individuals (Carusi & Jirotka 2009; Thomson et al 2005).

It is important to reiterate that this research is not actually focused on the individual CM's career as such, but in larger patterns of organising and the entanglement of financial actors. The transformation of ego-net data into interorganisational data effectively removes the individual. But where egos remain nodes in the network for analytic purposes, anonymising ego and firm alters effectively limits the possibility of individuals' becoming identifiable, even though it is no absolute guarantee that individuals cannot be re-identified (Rocher, Hendrickx & de Montjoye 2019).

Any data anonymisation effort should be consistently applied across the data set for both individuals and firms. Individuals can be coded using a simple ID-theme as done in Figures 3a and 3b, however, it may also be helpful to retain certain characteristics in the code; gender and age being obvious, although less useful ones. But it could be beneficial to retain an anonymised reference to the previous employer or the CMF in the code for the benefit of the researcher. Likewise, firm's names could be used for the analysis of interfirm ties (which excludes individuals), doing so may inadvertently allow individuals to be identified in the dataset. To avoid this firms should also be anonymised, for example, via a randomly generated code. It may, however, be more sensible to code firms according to specific characteristics, for example, according to their functional expertise: *Goldman Sachs* could be labelled 'Investment Bank A', *Linklaters* may be 'Law Firm 1' etc. The same applies to the discussion of educational attainment which may want to make reference to university status—Ivy League, Public Ivy, community college—because it may tell us something about the socio-economic status of individuals and the network opportunities that result from different universities. Anonymising individuals and firms limits the possibility for individuals to be traced in line with ethical-legal requirements set out by GDPR.

Gaining ethical approval for the study was complicated by the use of LinkedIn for triangulation purposes. Academic researchers are conscious about previous cases using 'social network' data has created outcomes that could potentially harm individuals. Zimmer's (2010) paper "But the data is already public" highlights these concerns by reviewing a study involving students'

Facebook profiles. Since then, the *Cambridge Analytica* scandal has intensified public discussions about what is permissible and what is not. Debates about the public *vis-a-vis* private nature of this data are complex and all but settled and concerns must be considered on a one-by-one basis, depending on the type of data used, the purpose and the origin of the data.

Social media is a broad category, it contains a large variety of services used for different purposes and that contains different data (see Bos et al 2009). While the use of Facebook data causes ample controversy (Zimmer & Kinder-Kurlanda 2017), it appears to be more permissible to analyse *Twitter* data, even when data collected is concerned with grieving individuals (Glasgow et al. 2016). And this makes sense given the nature of the data that is provided by these profiles and its use: Twitter amplifies free speech and personal opinion so that it can be heard globally, whilst Facebook profiles document the personal social life of an individual within a self-built social circle, i.e. his or her friends. Facebook is specific in the sense that it allows users to keep their profile private to safeguard their data and interactions, whereas Twitter is built on the premise of sharing opinions.

LinkedIn provides an alternative case yet again. It does not target naive individuals but more sophisticated *professionals* and this is reflected in the service it advertises: hire, market, sell and learn[3]. Whilst professionals may have ultimately different reasons for joining LinkedIn, the underlying principle is one of *discoverability* and *opportunity*. Discoverability refers to the intention of professional members to become visible beyond their traditional off-line network in real life: individuals connect to specific companies, follow influencers, and ask other users to promote skills or provide testimony of their achievements. The contents of their personal pages are carefully crafted and distinctly work-related in character (Papacharissi 2009). Opportunity considers the reason for individuals joining the network: to promote themselves to others; or as LinkedIn puts it, to "[c]reate economic opportunity for every member of the global workforce"[4]. People can and do advertise themselves to potential future employers or recruitment agents, and contact firms directly. In this view privacy is arguably of lesser concern to the users, not least because the context is professional, rather than social and it is more about being seen to be connected than actually exchanging information with connections (Davenport 2007). LinkedIn promotes connections over interactions and takes on the form of an online "online Rolodex" (Papacharissi 2009: 208). Any social interactions occur through direct communication via private messages or offline, outside of the publics' eyes. Of course, this do not provide us with a *carte blanche* to use data for whatever purpose, but it highlights that LinkedIn profiles are distinct to Facebook and other social media users. They are carefully crafted for a professional purpose and less social in character.

Requesting permission to conduct the research from LinkedIn was difficult not least because the lack of formal processes in place. To some extent this is intended by LinkedIn to limit its liability. However, I have received two confirmations that in this particular case it is acceptable to use LinkedIn for the intended purpose of triangulating data, but various caveats are attached to it and LinkedIn makes references to the Do's and Don'ts. A key stipulation is that data

[3] See https://about.linkedin.com/ and https://business.linkedin.com/# for additional information

[4] Quote retrieved from LinkedIn website https://about.linkedin.com/

triangulation must occur manually without using a computational approach including any software, scripts, or other means such as crawlers to scrape data.

In addition, I sought advice from my university's ethics and legal teams. Again, this was not a straightforward process and took considerable effort involving multiple clarifications and reaching out to different experts in the field. This was not necessarily because people disagreed on the use of such data for research purposes but because of the vague guidance by LinkedIn. Queries focused on the purpose of the research, the data use and subject anonymity. For this research LinkedIn profiles were 'consulted' alongside already existing biographies that are publicly available. The primary use of LinkedIn is not the discovery of data but assuring data accuracy and completeness. Where additional data is collected, it is ethically justifiable to do so as long as they correspond with already existing categories not least because of the emphasis on GDPR principles of purpose limitation and data minimisation (ICO 2018: 24ff.).

Consent is an additional issue that requires careful engagement with the purpose of the research and the private/public nature of the data. As argued by Wilson et al. (2012) it is insufficient to claim that all data is public to avoid consent. However, equally we must consider the difficulty of eliciting consent from all participants, a problem recognised given the distance between research subject and researcher (Zimmer & Kinder-Kurlanda 2017; Buchanan & Zimmer 2018). The question here is really whether or not consent is required after all. The answer to this may sound counter-intuitive; but, new GDPR guidelines stipulate that it is legitimate to process data without consent if it is in the public interest (see ICO 2018: 77ff.). UK Government guidance further stipulates that where consent is impractical due to the size of the dataset, researcher must be conducted in line with the terms and conditions and must protect indiviudals (Social Media Research Group, 2016).

Reflections on documents as source of network data

This paper discussed issues arising from collecting social network data from publicly available technical-legal documents. Whilst the context is specific, it provides some useful insight to researchers who seek to collect or construct network data from documents. Financial market activity is notoriously opaque and difficult to research, especially for industry outsiders. The use of documents mitigates this struggle to access information. Documenting is a prevalent feature of social, political and economic life and many of these provide data that may be of use to network research; contracts and policy documents may be particularly useful as they a published either comprehensively or regularly. The data collection described here results in networks that are comparable to interlocking directorates extracted from BoardEx or Orbis. However, the networks generated from documents also include employees with lees seniority, that is below the rank of directors. This additional information provides additional information on interfirm connectivity and knowledge transfer between organisations.

Sourcing network data from documents also has implication for the construction and subsequent analysis of ties. Data is collected retrospectively from documents which are not aimed at informing academic research and the researcher is not involved in collating the information. Therefore, the construction of ties and tie strength is informed by the empirical setting and the research question. An initial recording of ties in uniform format, for example, length of tenure, generates a dataset that can be easily manipulated before being analysed further. The discussion of strong, weak and negative ties illustrates one mechanism which could be applied. The comparison of different coding strategies highlights the impact even a basic recoding can have on the dataset. Whilst these will not be used in the final study, they may inform more advanced coding strategies such as the introduction of a decay function when analysing the data. The impact of applying more advanced coding strategies on the full network will be analysed and discussed in future publication.

Collecting network data from documents is not necessarily better or worse than other methods, but it is different. Documents do not suffer from recall bias which may impact on the quality of data gathered through questionnaires or interviews. Their content is fixed at the date of publication and can be revisited. Data quality is high, at least for those documents that have implications for public policy or commerce as they are internally scrutinised prior to publication (senior staff, legal or communications team) or externally (lawyers or professional services). Information provided may also be presented in a standardised format; for example, procurement contract would provide fixed information on the buyer and provider of goods and services, with key contact details and some descriptive justification to signal the expertise of the provider. Yet this fixity also poses problems to the researcher as the level of detail may differ and we cannot seek clarification from with human subjects to illicit additional information. To mitigate this issue, alternative sources of data, such as LinkedIn, have been located to triangulate data which produces legal and ethical challenges which must be carefully considered within the specific research context.

The data collection approach outlined in this paper provides guidance for researchers interested in or reliant on using documents as primary source for network data. Whilst the application focuses on non-participatory populations, it can also be used more broadly to elicit data suitable for network analysis from documents. Moreover, network data collected from documents could also be used as a baseline which may be used to collect additional data directly from participants where this is appropriate. A comparison of networks generated from documentary information versus participant-informed networks presents an opportunity for further research.

References

Bailey, K. D. (1994) *Methods of Social Research* (fourth edition). New York: The Free Press.

Baker, W. (2014) Making pipes, using pipes: How tie initiation, reciprocity, positive emotions and reputation create new organizational social capital. In D. Brass, G. Labianca, A. Mehra, D. Halgin, & S. P. Borgatti (Eds.), *Contemporary perspectives on organizational social networks* (pp. 57-72). Bradford, UK: Emerald Group Publishing.

Bavoso, V. (2017) Filling the Accountability Gap in Structured Finance Transactions: The Case for a Broader Fiduciary Obligation. *Columbia Journal of European Law*. 23(2): 369-400.

Beaverstock, J.V. & Hall, S.J.E. (2012) Competing for talent: global mobility, immigration and the City of London's labour market. *Cambridge Journal of Regions, Economy and Society*, 5(2): 271–288.

Beaverstock, J.V., Faulconbridge, J.R. & Hall, S.J.E. (2014) *The Globalization of Executive Search*. New York: Routledge.

Bison, I. (2014) Sequence as Network: An attempt to apply network analysis to sequence analysis. In Philippe Blanchard, Bühlmann, F. & Gauthier, J. (eds) *Advances in Sequence Analysis*. Cham: Springer.

Borgatti, S. P. & Everett, M. G. (2014) Networks containing negative ties. *Social Networks*, 38: 111-120.

Borgatti, S. P., Everett, M. G. & Johnson, J. C. (2013) *Analyzing Social Networks*. London, UK: Sage.

Bos, N., Karahalios, K., Musgrove-Chavez, M., Poole, E., S., Thomas, J. C. & Yardi, S. (2009) research ethics in the Facebook era: privacy, anonymity, and oversight. *Proceedings of Human Factors in Computing Systems,* 2009: 2767-2770.

Breure, A.S.H. & Heiberger, R.H. (2019) Reconstructing science networks from the past. Eponyms between malacological authors in the mid-19th century. *Journal of Hirstorical Network Research*, 3: 92-117.

Buchanan, E. A. & Zimmer, M. (2016) Internet Research Ethics. In Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Available at https://plato.stanford.edu/archives/fall2018/entries/ethics-internet-research/

Caccioli, F., Barucca, P. & Kobayashi, T. (2018) Network models of financial systemic risk: a review. *Journal of Computational Social Science* 1, 81–114.

Carusi, A. & Jirotka, M. (2009) From data archive to ethical labyrinth. *Qualitative Research*, *9*(3), 285–298.

Chernenko, S. (2017) The Front Men of Wall Street: The Role of CDO Collateral Managers in the CDO Boom and Bust. *Journal of Finance,* 72(5): 1893-1936.

Cohen, L., Manion, L. & Morrison K. (2007) *Research Methods in Education*. Abindon, UK: Routledge.

Crossley, N., Bellotti, E. , Edwards, G. , Koskinen, J. , Everett, M. & Tranmer, M. (2015) *Social Network Analysis for ego-nets*. London, UK: Sage.

Crossley, N., Edwards, G., Harries, E. & Stevenson, R. (2012) Covert social movement networks and the secrecy-efficiency trade off: The case of the UK suffragettes (1906–1914). *Social Networks*, 34(4): 634-644.

Davenport, T.H. (2007) LinkedIn Is Not a Social Network. *Harvard Business Review*, 2007. Available at: https://hbr.org/2007/11/linkedin-is-not-a-social-netwo.

Elo, K. (2018) Geospatial Social Networks of East German Opposition (1975-1989/90). *Journal of Historical network Research*, 2: 143-165.

Everett, M., Broccatelli, C., Borgatti, S.P. & Koskinen, J. (2018) Measuring knowledge and experience in two mode temporal networks. *Social Networks*, 55: 63-73.

FCIC (2011) *The Financial Crisis Inquiry Report*. Financial Crisis Inquiry Commission, Washington, D.C.: U.S. Government Printing Office.

Flick, U. (1992) Triangulation Revisited: Strategy of Validation or Alternative? *Journal for the Theory of Social Behaviour,* 22(2): 175-197.

Gai, P. & Kapadia, S. (2010) Contagion in Financial Networks. Bank of England Working Paper No. 383.

Glasgow K., Vitak J., Tausczik Y., Fink C. (2016) "With Your Help... We Begin to Heal": Social Media Expressions of Gratitude in the Aftermath of Disaster. In: Xu K., Reitter D., Lee D., Osgood N. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2016. Lecture Notes in Computer Science, vol 9708. Cham: Springer.

Godechot, O. (2008) "Hold-up" in Finance: The Conditions of Possibility for High Bonuses in the Financial Industry. *Revue française de sociologie*, 49(5): 95-123.

Granovetter, M. (1973) The Strength of Weak Ties. *American Journal of Sociology*, 78(6): 1360-1380.

Granovetter, M. (1974) *Getting a Job: A Study of Contacts and Careers*. Chicago, IL: University of Chicago Press.

Granovetter, M. (1985) Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology,* 91(3): 481-510.

Haldane, A. G. & May, R. (2011) Systemic risk in banking ecosystems. *Nature* 469, 351–355.

Heemskerk, E., Young, K., Takes, F.W., Cronin, B., Garcia-Bernardo, J., Henriksen, L.F., Winecoff, W.K., Popov, V. & Laurin-Lamothe, A. (2017) The promise and perils of using big data in the study of corporate networks: problems, diagnostics and fixes. *Global Networks*, 18(1): 3-32.

ICO (2018) Guide to the General Data Protection Regulation (GDPR). Information Commissioner's Office, Wilmslow, UK.

Jupiter High Grade CDO (2004) Jupiter High Grade CDO - Pitchbook. Available at: http://fcic-static.law.stanford.edu/cdn_media/fcic-docs/2004-12-00_Jupiter%20High-Grade_CDO%20Pitchbook.pdf

Labianca, G. & Brass, D. J. (2006) Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *Academy of Management Review*, 31(3): 564-614.

Author B

Marsden, P. V. (1990) Network Data and Measurement. *Annual Review of Sociology*, 16: 435-463.

Marsden, P. V. (2005) Recent Developments in Network Measurement. In Peter J. Carrington, John Scott, and Stanley Wasserman (eds.) *Models and Methods in Social Network Analysis*. New York, NY: Cambridge University Press.

Mizruchi, M.S. (1996) What do Interlocks do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates. *Annual Review of Sociology*, 22: 271-98.

Padgett, J. F. & Ansell, C. K. (1993) Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, 98(6): 1259-1319.

Papacharissi, Z. (2009) The virtual geographies of social networks: a comparative analysis of Facebook, LinkedIn and ASmallWorld. *New Media & Society*, 11(1&2): 199-220.

Perry, B.L., Pescosolido, B.A. & Borgatti, S.P. (2018) Egocentric Network Analysis: Foundations, Methods, and Models

Preda, A. (2002) Financial Knowledge, Documents, and the Structures of Financial Activities. *Journal of Contemporary Ethnography*, *31*(2): 207–239.

Prior, L. (2008) Repositioning Documents in Social Research. *Sociology*, 42(5): 821-836.

Reeves-Latour, M. & Morselli, C. (2017) B-id-rigging networks and state-corporate crime in the construction industry. *Social Networks*, 51: 158-170.

Reuters (2010) Greywolf Cap battles with bank over failed CDO. Reuters, May 6th 2010. Available at: https://www.reuters.com/article/greywolf-gs/greywolf-cap-battles-with-bank-over-failed-cdo-idUSN0610707520100506

Riles, A. (2006) *Documents: artifacts of modern knowledge*. Ann Arbor, MI: University of Michigan Press.

Rocher, L., Hendrickx, JM & de Montjoye, YA (2019) Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications,* 10(1): 1-9.

Roden, B. et al (this issue)

Rowlinson, M., Hassard, J. & Decker, S. (2014) Research Strategy in Organizational History: A Dialogue between Historical Theory and Organizational Theory. *Academy of Management Review*, 39(3): 250-274.

Scott, J. (1990) *A matter of record: Documentary sources in social research*. Cambridge, UK: Polity Press.

Scott, J. (1991) Social Network Analysis: A Handbook. London: Sage.

SEC (2010) Complaint: Goldman, Sachs & Co. and Fabrice Tourre. Available at: https://www.sec.gov/litigation/complaints/2010/comp-pr2010-59.pdf

SEC (2016) SEC Enforcement Actions: Addressing misconduct that let to or arose from the Financial Crisis. U.S. Securities and Exchange Commission. Available at: http://www.sec.gov/spotlight/enf-actions-fc.shtml

Shankar, K., Hakken, D., & Østerlund, C. (2017) Rethinking documents. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The Handbook of Science and Technology Studies* (pp. 59–86). Cambridge, MA: MIT Press.

Social Media Research Group (2016) Using social media for social research: An introduction. Government Social Research. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf

Stys, P., Verweijen. J., Muzuri, P. Muhindo, S. Vogel, C. & Koskinen, J. (2020) Brokering between (not so) overt and (not so) covert networks in conflict zones. *Global Crime*, 21(1): 74-110.

Suchman, M. S. (2003) The Contract as Social Artifact. *Law & Society Review*, 37(1): 91-142.

Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T. & Estabrooks, C. A. (2005) Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(1): 1-18.

Author (A)

US PSI (2011) Wall Street and the financial crisis: Anatomy of a financial collapse. US Permanent Subcommittee on Investigations, April 2011, Washington, D.C.

Useem, M. (1995) Reaching Corporate Elites. In, Rosanna Hertz & Imber, J.B. (eds) *Studying Elites Using Qualitative Methods*. Thousand Oaks, Sage, pp. 18-39.

Valeeva, D. Heemskerk, E.M. & Takes, F.W. (2000) The duality of firms and directors in board interlock networks: A relational event modeling approach. *Social Networks,* 62: 68-79.

van der Hulst, R. C. (2009) Introduction to Social Network Analysis (SNA) as an investigative tool. *Trends in Organised Crime*, 12: 101-121.

Wasserman, S. & Faust, K. (1994) *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Zimmer, M. (2010) “But the data is already public”: on the ethics of research in Facebook. *Ethics and Information Technology*, 12(4): 313-325.

Zimmer, M. & Kinder-Kurlanda, K. (2017) *Internet research ethics for the social age: New challenges, cases, and contexts.* New York, NY: Peter Lang Publishing.