Exploring Financial Relationships Using Probabilistic Topic Models (Demonstration Paper)

Louiqa Raschid University of Maryland Iouiqa@umiacs.umd.edu Zheng Xu University of Maryland xuzh@cs.umd.edu Elena Zotkina University of Maryland ezotkina@umiacs.umd.edu

ABSTRACT

Understanding relationships among financial entities can provide insight into the behavior of complex financial eco-systems. In this demonstration paper, we consider datasets of financial documents that describe the activity or role played by a financial institution (FI), typically with respect to a financial product or another financial entity. We develop community models based on financial institutions (FI) and their behavior or activity described by their roles (Role). Our models are based on an intuitive assumption that FIs will form communities, and FIs within a community are more likely to collaborate with other FIs in that community, and to play the same role, in other communities. Inspired by the Latent Dirichlet Allocation (LDA) and topic models, we develop several probabilistic financial community models and we use those models to identify interesting financial communities in two datasets.

Keywords

Latent Dirichlet allocation; generative probabilistic model; financial communities.

1. INTRODUCTION

There is a long and successful history of research in computational finance and financial engineering. The focus of much of this prior work included models for pricing or valuation of products; algorithms to make trading decisions; agents and simulations to understand the behavior of markets. In this research, we shift the focus to understand relationships between financial entities, and the roles that these entities play in a financial contract or product.

The relationships between financial entities are diverse and complex and may not be transparent. The most well understood relationship is of a parent entity with its subsidiaries or affiliates; these relationships are typically transparent. Other important transactional relationships such as a private ownership position or a repurchase agreement (repos) are typically hidden. There are indirect relationships

DSMM'17, May 14, 2017, Chicago, IL, USA

© 2017 ACM. ISBN 978-1-4503-5031-0/17/05...\$15.00

DOI: http://dx.doi.org/10.1145/3077240.3077247

when financial entities participate on a financial contract; an example is the complex supply chain required to create a mortgage backed securities contract. There are also business relationships between competitors, or contextual relationships, e.g., plaintiff and defendant in litigation around a financial product.

In this demonstration paper, we consider two datasets of financial documents (FD) that are filed with the Securities and Exchange Commission (SEC). One dataset contains the prospectuses for residential mortgage backed securities (resMBS). These resMBS prospectuses typically are required to provide precise details about the role played by each participating financial entity. We can expect the same roles to be repeated across the documents, and this dataset will be relatively complete with respect to knowledge about the entities and their roles. A second dataset includes annual and quarterly SEC filings (10-K and 10-Q) that discuss the activities of a financial entity. Each entity can decide on the activity to be reported, and these relationships may be complex and incompletely described in the documents. In each case, we extract the following triples:

(FD, Role, FI = Mentioned entity).

We assume that the mentioned entity (FI) is playing the corresponding role with respect to the resMBS contract described in the FD, in the first dataset. In the second dataset, we assume that the mentioned entity is playing the role with with respect to the entity that is filing the FD (10-K or 10-Q).

Our models are motivated by the intuition that financial entities will form communities, and that an entity in a community is more likely to collaborate, with other entities in that community, on other products. An entity will also continue to play the same role, in other financial contracts or other communities. Inspired by the Latent Dirichlet Allocation (LDA) and topic models, we develop several probabilistic financial community models and we use those models to identify interesting financial communities. Each financial document (FD) is represented as a bag of words (phrases), where the words are the mentioned entities (FI) and the roles (Role).

We develop the following models: 1. FI-C: each FD is a bag of FIs (words); 2. Role-FI-C: each FD is a bag of Role-FI pairs (words);

Our observations are as follows: 1. Some communities reflect known relationships while others may provide some financial insight into the behavior of some financial products. 2. The topics uncovered from FI-C and Role-FI-C models show that the role played by an entity can be very signifi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

cant. 3. From resMBS Role-FI-C topics, we isolated topics where financial entities that were bad players prior to the 2008 US financial crisis were dominant and played multiple roles. These may be considered *toxic topics* that are of interest to regulators.

There has been prior work on extracting financial entities and relationships from documents [3, 4] To our knowledge, our work reported in [10] is the first to apply probabilistic graphical models to understand financial communities. We continue that research with this demonstration paper reporting on experiments to explore financial communities.

2. DATASETS

We consider two datasets, resMBS and FEIIIY2; both are public documents filed with the Securities and Exchange Commission (SEC) [8]. resMBS includes 5000+ prospectuses for residential mortgage backed securities that were issued by private labels and filed with the SEC between 2002 and 2008. FEIIIY2 includes 150+ 10-K filings and 500+ 10-Q filings from 27 financial entities filed between 2011 and 2016.

We briefly discuss the extraction process to create the datasets; details are available in [3, 9]. The extraction pipeline is developed using the rule-based algebraic information extraction system, SystemT [5]. Dict NER [9] is a special purpose Named Entity Recognizer that is tuned to extract financial institution (FI) names. FI names are typically composed of a root, which is usually unique, and a suffix which is drawn from a small corpus of suffix terms. Dict NER utilizes both a root dictionary and a suffix dictionary to recognize FI names. Rank ER [9] performs entity resolution on the extracted FI name and maps each FI name to a corpus of standardized FI names obtained from multiple sources including the ABSNet portal [1] and the National Information Center of the Federal Reserve System [7]. A Role Extraction module uses keyword matching to extract roles such as issuer, depositor, sponsor, etc. Role keywords were defined for the two datasets. A Role Participant Matching module pairs a role with one or more FI names.

Figure 1 illustrates the summary section of a resMBS prospectus for a mortgage backed security. Example FI names in this summary are Wachovia Bank, National City, HSBC Bank, etc. We can also extract the Role played by the FIs, e.g., depositor, issuing entity, seller, sponsor, originator, servicer, trustee, etc. Consider the three columns in the lower part of the figure; this includes the Role, the extracted name of the mentioned entity, and the matching standardized FI name (determined after entity resolution). For example, Wachovia, identified as FI380, plays the role of depositor, issuing entity, seller and sponsor, for this exemplar FD. Similarly, National City Bank, identified as FI263, plays the role of originator and servicer.

Details about the quality of the FI mentions and Role Participant matching are in [3, 6, 9]. We note that the quality of the extracted triples dataset is very good for resMBS. For FEIIIY2, the performance of the Role Participant Matching module requires additional tuning. We only used a subset of the FEIIIY2 that appeared to be of higher quality.

Summary statistics for resMBS FI-C, resMBS Role-FI-C and FEIIIY2 (for FI-C and Role-FI-C) are in Table 1.

	resMBS	resMBS	FEIIIY2
	FI-C	Role-FI-C	both
Count of documents	3146	4472	562
Count of distinct FIs	96	96	214
Count of FI occurrences	36945		7155
Count of distinct Roles		33	9
Distinct (Role_FI) pairs		267	409
Count of (Role_FI) occ.		41075	7155

Table 1: Summary statistics for the resMBS and FEIIIY2 financial documents.

3. FINANCIAL COMMUNITY MODELS

Latent Dirichlet Allocation (LDA) [2] is a generative probabilistic model for collections of discrete data (documents). LDA represents each document as a random mix over latent topics, where each topic is characterized by a distribution over words. Given the hyper-parameters α and β , the probability of a document with N words is

$$p(\mathbf{w}|\alpha,\beta) = \int p(\theta|\alpha) (\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta) p(w_n|z_n,\beta)) d\theta \quad (1)$$

where $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$ is a set of N words, θ is a topic mixture sampled from a Dirichlet distribution parameterized by α , z_n represents a topic sampled from a multinomial distribution conditioned on θ , and each word w_n is sampled from a multinomial distribution conditioned on z_n and parameterized by β . LDA is a three-level hierarchical Bayesian model. The parameters α and β are corpus level parameters, assumed to be sampled once in the process of generating a corpus. The variables θ are document-level variables, sampled once per document. Finally, the variables z_n and w_n are word-level variables, sampled once for each word in each document.

We use FD to denote a financial document, FI to denote a financial institution, and Role-FI to denote a pair comprising a role and an FI. We build three models, FI-C, Role-FI-C and UNION-C. In FI-C, each FD is a bag of FIs (words), corresponding to a random mix over latent probabilistic FI communities, where each probabilistic FI community is characterized by a distribution over FIs. In Role-FI-C, each FD is a bag of Role-FI pairs (words), and each probabilistic Role-FI community is characterized by a distribution over Role-FI pairs. UNION-C represents each FD as the union of FIs and Role-FI pairs; similarly, each community is a distribution over both FIs and Role-FI pairs.

The models were implemented using the Python *sklearn* toolkit. ¹ We evaluated the quality of the communities (topics) using a coherence metric and the explanation and generative ability of our models over test documents using a perplexity metric; details are presented in [10].

4. DEMONSTRATION RESULTS

Given that this was a small dataset, we set the number of topics to 30. Using the probability distribution of words across all the topics, we used a threshold of 0.05 to filter out words from topics. We then chose a threshold probability of 0.10 to determine when a word was *significant* to the topic.

 $[\]label{eq:linear} ^{1} http://scikit-learn.org/stable/modules/generated/sklearn. decomposition.LatentDirichletAllocation.html$

	Summary section of WACHOVIA_1393388_0001068238-07-00	0417			
Title of Series	Wachovia Mortgage Loan Trust, LLC Mortgage Pa Certificates, Series 2007-A.	iss-Through			
Depositor	Wachovia Mortgage Loan Trust, LLC.				
Issuing entity	Wachovia Mortgage Loan Trust, LLC Series 2007	'-A Trust.			
Seller and Sponsor	Wachovia Bank, National Association.				
Originators	National City Mortgage, Fifth Third Mortgage Mortgage, Inc. and Wells Fargo Bank, N.A.	Company, SunTrust			
Servicers	National City Mortgage Co., or National City, Company, SunTrust Mortgage, Inc. and Wells Fa	Fifth Third Mortgage Irgo Bank, N.A.			
Trustee	HSBC Bank USA, National Association.				
Master Servicer and Certificate Administrator					
	Extracted tuples:				
(Role;	Dict-NER Participant;	Rank-ER FI-ID)			
(Depositor;	Wachovia Mortgage Loan Trust, LLC;	FI380:Wachovia),			
(Issuing Entity;	tv: Wachovia Mortgage Loan Trust, LLC Series 2007-A; FI380:Wachovia).				
(Seller;	r; Wachovia Bank, National Association; FI380:Wachovia).				
(Sponsor;	onsor; Wachovia Bank, National Association; FI380:Wachovia).				
(Originators;	iginators; National City Mortgage; FI263:National City),				
(Originators;	iginators; Fifth Third Mortgage Company; FI145:Fifth Third Bank),				
(Originators;	iginators; SunTrust Mortgage, Inc.; FI345:SunTrust),				
(Servicers;	vicers; National City Mortgage Co.; FI263:National City),				
(Servicers;	cers; National City; FI263:National City),				
(Servicers;	Fifth Third Mortgage Company; FI145:Fifth Third Bank)				
(Trustee;	HSBC Bank Association; FI183:HSBC),				
(Master Servicer;	ster Servicer; U.S. Bank National Association; FI363:U.S. Bank),				

Figure 1: Summary section of an example prospectus and the extracted 3-tuples.

We note that further experiments are needed to determine a good set of parameters.

Topics that are composed of a single significant word or two words were considered to be simple topics and we do not explore them in this discussion. A visualization of more complex topics, each of which contains at least three significant words is available at the following site: ²

4.1 resMBS Communities

Figure 2 shows some resMBS FI-C topics. The topic identifier is on the left and the standardized FI name is on the right. The width of the edge corresponds to the probability of the word in that topic.

- We note that each topic is complex and contains an average of 5 FI names. The FI Wells Fargo is seen to be significant across multiple topics; the edge width also shows that it plays a very prominent role in these topics. Deutsche Bank is the next most significant FI.
- Not surprisingly, most topics correspond to known financial relationships. For example, in Topic 4, Aurora is a subsidiary of Lehman Brothers. Similarly, in Topic 7, EMC Mortgage is a subsidiary of Bear Stearns.
- The following topics reflect financial communities that do not share a known parent or affiliate relationship: Topic 6: (Deutsche Bank, IndyMac, Wells Fargo); Topic 8: (Cendant, Chase Manhattan, Wells Fargo); Such topics may be of interest to a financial analyst. For example, they may want to determine the number of prospectuses issued by some community, or they may want to study the performance of securities associated with prospectuses issued by some community.

Figure 3 shows selected resMBS Role-FI-C topics; recall that for these communities, the words are (Role, FI) pairs. For ease of visualization we do not show the role labels. What is notable is that the topics are very different compared to the FI-C topics of Figure 2. This reflects that the role played by the FI is indeed very significant.

- Our first observation is that Countrywide Securities Corporation and Countrywide Home Loans play a significant role in multiple Role-FI-C topics. This is in contrast to Figure 2 where only Wells Fargo is seen to be significant across multiple FI-C topics. An explanation is that Countrywide Securities Corporation and Countrywide Home Loans may be consistently playing the same role across multiple resMBS contracts. As a result, the corresponding (Role, FI) pairs become more significant in the Role-FI-C model. In contrast, an FI that has multiple roles across contracts may find its corresponding (Role, FI) pairs becoming less significant.
- Topics 6, 7 and 10 are of particular interest. Topic 6 is associated with IndyMac, topic 7 is associated with First Horizon Home Loan Corp and topic 10 is associated with Ameriquest. These three FIs, Ameriquest and First Horizon and IndyMac, all issued sub-prime mortgages that played a prominent role in the 2008 US financial crisis. Ameriquest failed in 2007 and IndyMac failed in 2008.

Figure 4 shows these resMBS Role-FI-C topics 6, 7 and 10, (from Figure 3) with their role labels displayed. We can see that each of these communities is dominated by one of the FIs. This FI notably plays multiple roles in the community.

²https://dsfin.umiacs.umd.edu/topicmodels/

F1382-WASHINGTON MUTUAL Topic 0 F1306-DLI MORTGAGE CAPITAL CORP F1211,P. MORGAN CHASE F1303-CREDIT SUISS SECUTITES (USA) LLC F1389-WILSHIRE CREDIT CORPORATION F16-AAMES CAPITAL CORPORATION Topic 1 F1303-SELECT PORTFOLIO SERVICING, INC. F1314RENAISSANCE MORTGAGE ACCEPTANCE F1314- Topic 2 F1389WILSHING CORPORATION F1314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 F1314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 F1314RENAISSANCE MORTGAGE HOME LOAN F1388			
Topic 0 FI10E-DLI MORTGAGE CAPITAL CORP FI211FAIRBANKS CAPITAL CORPORATION FI211FAIRBANKS CAPITAL CORPORATION FI6-AAMES CAPITAL CORPORATION FI6-AAMES CAPITAL CORPORATION FI6-AAMES CAPITAL CORPORATION FI6-AAMES CAPITAL CORPORATION FI6-AAMES CAPITAL CORPORATION FI314RENAISSANCE MORTGAGE ACCEPTANCE FI314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111DELTA FUNDING CORPORATION FI358THORNBURG MORTGAGE HOME LOAN Topic 2 FI384WELLS FARGO Topic 6 FI138DEUTSCHE BANK Topic 6 FI138DEUTSCHE BANK Topic 6 FI138DEUTSCHE BANK FI226LITION LOAN SERVICING LP FI226LITION ALARTAN BANK Topic 4 FI223LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI30-BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI226CITIMORTGAGE FI226CITIMORTGAGE FI226CITIMORTGAGE FI226CITIMORTGAGE FI226CITIMORTGAGE FI226CITIMORTGAGE FI226-DITIONAL INC. FI219GMAC		FI382WASHINGTON MUTUAL	
Injut 0 FI140-FAIRBANKS CAPITAL CORP FI103-CREDIT SUISSE SECURITIES (USA) LLC FI389-WILSHIRE CREDIT CORPORATION FI6-AAMES CAPITAL CORPORATION Topic 1 FI380-SELECT PORTFOLIO SERVICING, INC. FI314-RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111-DEITA FUNDING CORPORATION FI358-THORNBURG MORTGAGE ACCEPTANCE Topic 6 FI131-DEUTSCHE BANK Topic 6 FI131-DEUTSCHE BANK Topic 7 FI26-LITION LOAN SERVICING IP 04CREDIT-BASED ASSET SERVICING AND SECURITIES FI26-LITION LOAN SERVICING IP 04CREDIT-BASED ASSET SERVICING AND SECURITIES FI26-LITION LOAN SERVICING IP 04CREDIT-BASED ASSET SERVICING AND SECURITIES FI26-LITION LOAN SERVICING IP 04CREDIT-BASED ASSET SERVICING AND SECURITIES FI26-LITION LOAN SERVICING IP 04CREDIT-BASED ASSET SERVICING AND SECURITIES FI272-CENDANT MORTGAGE FI34-AURORA LOAN SERVICIES LLC Topic 7 FI30-BEAR STEARNS & CO. INC. FI129-EMC MORTGAGE FI32-CITIMONTGAGE FI278-NOMURA SECURITIES INTERNATIONAL INC. FI278-NOMURA SECURITIES INTERNATIONAL INC. <td>Tonic 0</td> <td>FI106DLI MORTGAGE CAPITAL INC.</td> <td></td>	Tonic 0	FI106DLI MORTGAGE CAPITAL INC.	
H1103CREDIT SUISSE SECURITIES (USA) LLC F1303WILSHIRE CREDIT CORPORATION F16-AAMES CAPITAL CORPORATION F16-AAMES CAPITAL CORPORATION Topic 1 F130SELECT PORTFOLIO SERVICING, INC. F1314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 F1111-DETLA FUNDING CORPORATION F1325THORNBURG MORTGAGE ACCEPTANCE Topic 3 F1111-DETLA FUNDING CORPORATION F1358THORNBURG MORTGAGE HOME LOAN Topic 6 F1134RENAISSANCE MORTGAGE HOME LOAN Topic 7 F1384STRUCTURE BANK F1226UTTON LOAN SERVICING LP OdCREDIT-BASED ASSET SERVICING AND SECURITIZATION F1224UTTON LOAN SERVICING LP OdCREDIT-BASED ASSET SERVICING AND SECURITIZATION F1224UTTON LOAN SERVICING LP F1224UTTON LOAN SERVICING LP F1224UTON LOAN SERVICING LP F1224UTON LOAN SERVICING LP F1224UTAN MORTGAGE F1224UTAN MORTGAGE F1224UTAN LOAN SERVICES LLC Topic 7 F150-BEAR STEARNS & CO. INC. F1121EMC MORTGAGE F122UTIMORTGAGE F123-NOMURA SECURITIES INTONAL INC. F121GMAC	Topic o	FI140FAIRBANKS CAPITAL CORP	
FI389WILSHIRE CREDIT CORPORATION FIG-AAMES CAPITAL CORPORATION Topic 1 FI330-SELECT PORTFOLIO SERVICING, INC. FI314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111-DELTA FUNDING CORPORATION FI358THORNBURG MORTGAGE HOME LOAN Topic 2 FI384WELLS FARGO Topic 6 F113DEUTSCHE BANK Topic 6 F113DEUTSCHE BANK Topic 7 FI38-HSBC FI226LITION LOAN SERVICING LP D4CREDIT-BASED ASSET SERVICING AND SECURITIZATION F1226LITION LOAN SERVICING LP D4CREDIT-BASED ASSET SERVICING AND SECURITIZATION F1226LITION LOAN SERVICING LP D4CREDIT-BASED ASSET SERVICING AND SECURITIZATION F1226LITION LOAN SERVICING LP D4CREDIT-BASED ASSET SERVICING AND SECURITIZATION F124CHASE MANHATTAN BANK Topic 4 F1223LEHMAN BROTHERS F134AURORA LOAN SERVICES LLC Topic 7 FI50-BEAR STEARNS & CO. INC. F1119EMC MORTGAGE F128CITIMORTGAGE F128CITIMORTGAGE F128NOMURA SECURITIES INTERNATIONAL INC. F1278-NOMURA SECURITIES INTERNATIONAL INC.		FI103CREDIT SUISSE SECURITIES (USA) II C	
FIG-AAMES CAPITAL CORPORATION Topic 1 Fi330SELECT PORTFOLIO SERVICING, INC. FI285OCWEN FI314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111-DEITA FUNDING CORPORATION FI384WELLS FARGO Topic 2 Topic 6 Topic 7 Topic 8 FI226-LITTON LOAN SERVICING IP O4CREDIT-BASED ASSET SERVICING AND SECURITIES FI226-LITTON LOAN SERVICING IP O4CREDIT-BASED ASSET SERVICING AND SECURITIES FI226-LITTON LOAN SERVICING IP OFIC 8 FI226-LITTON LOAN SERVICING IP OFIC 8 FI226-LITTON MORTGAGE FI226-LITTON MORTGAGE FI226-LEHMAN BROTHERS FI226-LITMORTGAGE FI226-LITMORTGAGE <td></td> <td>FI389WILSHIRE CREDIT CORPORATION</td> <td></td>		FI389WILSHIRE CREDIT CORPORATION	
Topic 1 FI330-SELECT PORTFOLIO SERVICING, INC. FI285OCWEN FI314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111DEITA FUNDING CORPORATION FI358THORNBURG MORTGAGE HOME LOAN Topic 2 FI384WELLS FARGO Topic 6 FI113DEUTSCHE BANK FI205INDYMAC		EIGAAMES CAPITAL CORPORATION	
Fi285-OCWEN Fi285-OCWEN Fi285-OCWEN Fi285-OCWEN Fi285-OCWEN Fi285-THORNBURG MORTGAGE ACCEPTANCE Topic 3 Topic 2 Topic 6 Fi205INDYMAC Topic 9 Fi226UTON LOAN SERVICING LP OURESED ASSET SERVICING AND SECURITIZATION Fi226UTON LOAN SERVICING LP FI22CENDANT MORTGAGE FI226UTON LOAN SERVICIES LLC Topic 7 FI36ULEMMAN SERVICIES LLC Topic 7 FI32ULEMMAN SERVICIES LLC Topic 5 FI84CITIMORTGAGE	Taula 1	FI330SELECT PORTFOLIO SERVICING, INC.	
FI314RENAISSANCE MORTGAGE ACCEPTANCE Topic 3 FI111DELTA FUNDING CORPORATION FI358THORNBURG MORTGAGE HOME LOAN Topic 2 FI384WELLS FARGO Topic 6 F113DEUTSCHE BANK Topic 7 FI383-HSBC FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OH-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI226LITION LOAN SERVICING LP OTOPIC 4 FI223LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FIS0-BEAR STEARNS & CO. INC. FI129EMC MORTGAGE FI278-NOMURA SECURITIES INTERNATIONAL INC. FI278-NOMURA SECURITIES INTERNATIONAL INC. FI278-NOMURA SECURITIES INTERNATIONAL INC.	Topic I	FI285OCWEN	
Topic 3 FI11-DELTA FUNDING COOPORATION Topic 2 FI384WELLS FARGO Topic 6 FI113DEUTSCHE BANK FI205INDYMAC FI205INDYM	E12	14. DENAISSANCE MORTGAGE ACCEPTANCE	
HISS8-THORNBURG MORTGAGE HOME LOAN Topic 2 FI384WELLS FARGO Topic 6 FI205INDYMAC FI207-CENDANT MORTGAGE FI205INDYMAC FI205INDYMAC FI205INDYMAC FI205INDYMAC FI205INDYMAC FI205INDYMAC FI205INDYMAC FI206-INDY FI206-INDY FI207-GMAC FI208-INDY FI208-INDY FI208-INDY FI208-INDY FI208-INDY FI208-INDY	Tonic 3	FI111DELTA FUNDING CORPORATION	
Topic 2 Fi384WELLS FARGO Topic 6 F113-DEUTSCHE BANK Topic 6 F1205	FI	358THORNBURG MORTGAGE HOME LOAN	
Topic 2 Topic 2 Topic 6 FI133-DEUTSCHE BANK FI205INDYMAC FI133HSBC FI205INDYMAC FI133HSBC FI205LITION LOAN SERVICING LP FI205LITION LITION LI			
Topic 2 Topic 2 Topic 6 FI134WELLS FARGO FI205INDYMAC			
Topic 2 Topic 6 F113DEUTSCHE BANK F1205INDYMAC F1133HSBC F1236ITFON LOAN SERVICING LP F1226ITFON LOAN SERVICING LP F1226ITFON LOAN SERVICING LP F1226		FI384WELLS FARGO	
Topic 6 FI133-DEUTSCHE BANK FI205INDYMAC FI183HSBC FI226LITTON LOAN SERVICING LP FI226LITTON LOAN SERVICES LLC Topic 7 FI30BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI226NOMURA SECURITIES INTERNATIONAL INC. FI276NOMURA SECURITIES INTERNATIONAL INC.	Topic 2		
Topic 6 FI13DEUTSCHE BANK FI205INDYMAC Topic 9 FI3-AES FI25UITION LOAN SERVICING AND SECURTIZATION FI26LITION LOAN SERVICING AND SECURTIZATION FI34STRUCTURED ASSET FI72CENDANT MORTGAGE FI78CHASE MANIATTAN BANK Topic 4 FI223LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI50-BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE FI32CITIMORTGAGE			
Topic 6 F113-H50C F123-H50C F1			
FI205INDYMAC FI205INDYMAC FI183HSBC FI262UTTON LOAN SERVICING LP IO4CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI24STRUCTURED ASSET FI72CENDANT MORTGAGE FI78CHASE MANHATTAN BANK Topic 4 FI223LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI50BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE	Topic 6	FILIS-DEUTSCHE BANK	
Topic 9 FI133-HSBC FI236-LITTON LOAN SERVICING LP IO4-CREDIT-BASED ASSET SERVICING AND SECURTIZATION IOPIC 8 FI244-STRUCTURED ASSET FI78-CHASE MANHATTAN BANK TOPIC 4 FI230-LEHMAN BROTHERS FI34-AURORA LOAN SERVICES LLC Topic 7 FI30-BEAR STEARNS & CO. INC. FI119-EMC MORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI278-NOMURA SECURITES INTERNATIONAL INC. FI219-GMAC		FI205INDYMAC	
100-CE SECURITIES F1226-UTON LOAN SERVICING LP 104-CREDIT-BASED ASSET SERVICING AND SECURITIZATION F1344-STRUCTURED ASSET F172-CENDANT MORTGAGE F178-CHASE MANHATTAN BANK F1363-U.S. BANK Topic 4 F122-LEHMAN BROTHERS F134-AURORA LOAN SERVICES LLC Topic 7 F150-BEAR STEARNS & CO. INC. F1119-EMC MORTGAGE F182-CITIMORTGAGE F182-CITIMORTGAGE F1278-NOMURA SECURITIES INTERNATIONAL INC. F1171-GMAC	Tonic 9	FI183HSBC	
1926-UITON LOAN SERVICING LP 104-CREDIT-BASED ASSET SERVICING AND SECURITIZATION FI344-STRUCTURED ASSET FI72-CENDANT MORTGAGE FI78-CHASE MANHATTAN BANK FI78-CHASE MANHATTAN BANK FI78-CHASE MANHATTAN BANK FI78-CHASE MANHATTAN BANK FI72-CENDANT MORTGAGE FI34-AURORA LOAN SERVICES LLC Topic 7 FI50-BEAR STEARNS & CO. INC. FI119-EMC MORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI82-CITIMORTGAGE FI82-CITIGROUP FI278-NOMURA SECURITIES INTERNATIONAL INC. FI171-GMAC	. opie s	FI9ACE SECURITIES	
FI34STRUCTURED ASSET FI34STRUCTURED ASSET FI72CENDANT MORTGAGE FI78CHASE MANHATTAN BANK FI363U.S. BANK Topic 4 FI323LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI30BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE	INA CREDIT RA	FI226LITTON LOAN SERVICING LP	
Topic 4 FI72CENDANT MORTGAGE FI78CHASE MANHATTAN BANK FI363U.S. BANK FI363U.S. BANK FI32LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI30BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC	Topic 8	SED ASSET SERVICING AND SECORHIZATION	
FI72CENDANT MORTGAGE FI78CHASE MANIAITTAN BANK FI363U.S. BANK FI363U.S. BANK FI363U.S. BANK FI323LEHMAN BROTHERS FI34AURORA LOAN SERVICES LLC Topic 7 FI50BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE FI82CITIMORTGAGE FI82CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC		FI344STRUCTURED ASSET	
FI78CHASE MANHATTAN BANK FI363-U.S. BANK Topic 4 FI34AURORA LOAN SERVICES LLC Topic 7 FI30BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CTIMORTGAGE FI82CTIMORTGAGE Topic 5 FI82CTIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC		FI72CENDANT MORTGAGE	
Topic 4 FI223-LEHMAN BROTHERS FI34-AURORA LOAN SERVICES LLC Topic 7 FI50-BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE Topic 5 FI84CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC		FI78CHASE MANHATTAN BANK	
FI34AURORA LOAN SERVICES LLC FI34AURORA LOAN SERVICES LLC Topic 7 FI50BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE Topic 5 FI84CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC	Topic 4	FI223I FHMAN BROTHERS	
FI34-AURORA IOAN SERVICES LLC Topic 7 FI50-BEAR STEARNS & CO. INC. FI119EMC MORTGAGE FI82CITIMORTGAGE Topic 5 FI84CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC			
Topic 5 FI32CITIMORTGAGE FI32CITIMORTGAGE Topic 5 FI32CITIMORTGAGE FI32CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC.	Topic 7	FI34AURORA LOAN SERVICES LLC	
FI119EMC MORTGAGE FI82CITIMORTGAGE Topic 5 FI84CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI271GMAC	Topic 7	FI50BEAR STEARNS & CO. INC.	
FI82CITIMORTGAGE Topic 5 FI84CITIGROUP FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC		FI119EMC MORTGAGE	
Topic 5 Fi84CITIGROUP Fi278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC		FI82CITIMORTGAGE	
FI278NOMURA SECURITIES INTERNATIONAL INC. FI171GMAC	Topic 5	FI84CITIGROUP	
FI171GMAC	FI278-	-NOMURA SECURITIES INTERNATIONAL INC.	
		FI171GMAC	_

Figure 2: resMBS: FI Communities

It was well understood that prior to the financial crisis, there were known bad players that were issuers of resMBS securities contracts, originators of the sub-prime mortgages, or otherwise played a significant role in some toxic financial products. Our observations from Figure 4 and the resMBS Role-FI-C model may suggest that communities where an FI is significant, and where that same FI plays multiple roles, may be worth further investigation by financial analysts and financial regulators.

4.2 **FEIIIY2** Communities

We briefly comment on some sample FEIIIY2 communities. We observe that the role is indeed significant and as a result, the FI-C topics of Figure 5 and the Role-FI-C topics of Figure 6 show many differences.

For example, both BNY and U.S. Bancorp are significant in Figure 5 but are less significant in Figure 6. In contrast, State Street becomes very significant in Figure 6, where it plays the roles of counterparty and agent in topic 2. Similarly, Charles Schwab becomes very significant in Figure 6 where it plays the role of trustee on topic 0.

5. CONCLUSION AND FUTURE WORK

This demonstration paper used multiple probabilistic topic models to uncover communities in two datasets. The topics uncovered from FI-C and Role-FI-C models show that the role played by an entity can be very significant. From resMBS Role-FI-C topics, we isolated topics where financial entities that were bad players prior to the 2008 US financial crisis played multiple roles. In future work, we will further refine our models and explore the relevance of topics. We will do this by extending our datasets to include both relevant financial performance indicators as well as sentences from the documents that provide information about signifi-



Figure 3: resMBS: Role-FI Communities (with Role not displayed).

cant financial entities and their roles.

6. ACKNOWLEDGMENTS

This research was partially supported by NIST award 70NANB15H194 and NSF grants CNS1305368 and DBI1147144.



Figure 4: resMBS: Role-FI Communities (with Role displayed).

	FI66COUNTRYWIDE
	FI26BANK OF AMERICA
Topic 1	FI20ASSURED GUARANTY
	FI164NEW YORK
Tonic 2	
. opio 1	FI35BNY
Tania	
Topic o	FI226U.S. BANCORP
	FI82FANNIE MAE
	FI54CHARLES SCHWAB CORPORATION
Topic 3	FI207SUNTRUST
	FI58COLONIAL BANCGROUP
Taular	FI75DISCOVER
Topic 5	FIZZ5TRAVELERS
	FI79EQUITY SECURITIES
Topic 4	FI98GINNIE MAE
	FI85FHA
Topic 6	FI30BB&T
Topico	FI199STATE STREET
	FI96GENERAL ELECTRIC
Topic 7	FI56CITIGROUP
	FI144MORGAN STANLEY
Tania 0	FI102GOLDMAN SACHS
TOPIC 8	FI114IB FINANCE
	FI113-HUNTINGTON BANCSHARES
	EI119_ITA
	FI155NASDAQ
	FI103GREENPOINT
Topic 9	EI86EHI B
Topic S	FI94FRB
	FI192SAVINGS BANK

Figure 5: FEIIIY2: FI Communities.

7. **REFERENCES**

- [1] Absnet. http://www.absnet.net/ABSNet.
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.
- [3] D. Burdick, S. De, L. Raschid, M. Shao, Z. Xu, and E. Zotkina. resMBS: Constructing a financial supply chain graph from financial prospecti. In *SIGMOD DSMM*. ACM, 2016.
- [4] D. Burdick, M. A. Hernández, H. Ho, G. Koutrika, R. Krishnamurthy, L. Popa, I. Stanoi, S. Vaithyanathan, and S. R. Das. Extracting, linking and integrating data from public sources: A financial case study. *IEEE Data Eng. Bull.*, 2011.
- [5] L. Chiticariu, R. Krishnamurthy, Y. Li, S. Raghavan, F. R. Reiss, and S. Vaithyanathan. SystemT: an algebraic approach to declarative information extraction. In ACL, 2010.
- [6] Financial entity identification and information integration data challenges. https://ir.nist.gov/dsfin.
- [7] National information center. https: //www.ffiec.gov/nicpubweb/nicweb/NicHome.aspx.
- [8] Securities and exchange commission. https: //www.sec.gov/edgar/searchedgar/webusers.htm.
- [9] Z. Xu, D. Burdick, and L. Raschid. Exploiting lists of names for named entity identification of financial institutions from unstructured documents. arXiv preprint arXiv:1602.04427, 2016.
- [10] Z. Xu and L. Raschid. Probabilistic financial community models with latent dirichlet allocation for financial supply chains. In *SIGMOD DSMM*. ACM, 2016.



Figure 6: FEIIIY2: Role-FI Communities (with Role displayed).