

# Building a cross-generation social network for Jewish sages

Avital Zadok

## Abstract

## Background

The Mishnah is a comprehensive anthology of Jewish halachah which developed in Jewish study halls between the first century C.E. and start of the third century C.E. (רוזן-צבי, 2018). The Tosefta is an anthology of laws supplementing the Mishnah which was compiled following the pattern of the Mishnah and, to a great extent, follows its laws (אלבק, 1944; בלברג, 2013).

Throughout the ensuing eras, many biographical works were authored on august Rabbinical figures, beginning with the Rabbis of the Tanaic era, continuing with the Amoraic sages and culminating with the era of the Geonim, in manuscripts dating back as far as *Igeret Rav Sherira Gaon* [Letter of Rabbi Sherira Gaon] and continuing until the present day. Some of the more comprehensive recent biographical works include *Toldot HaTanaim veHaAmoraim* [History of the Tanaim and the Amoraim] (הימן, 1987); *Mavo LaMishnah* [Introduction to the Mishnah] (אלבק, 1944); *Mavo LaTalmudim* [Introduction to the Talmuds] (אלבק, 1969); *Encyclopedia LeChachmei HaTalmud VeHageonim* [Encyclopedia of Sages of the Talmud and the Geonim] (מרגליות, 1995); and *Atlas Eitz Chaim, Tanaim VeAmoraim* Vol. 1, 2 [Tree of Life Atlas, Tanaim and Amoraim] (הלפרין, 1980).

A partial summary of these studies can be found in the preface to the monumental bibliographical project “*Otsar HaDmuyot BaSifrut HaRabbanit*” [Treasure of Personalities in Rabbinical Literature] which was published by Bonayich Educational Services (בונייך שירותי חינוך, 2005). These information banks offer a comprehensive database that encompasses a list of some 3,000 names found in Rabbinical literature from the Mishnaic and Talmudic eras, along with pertinent data regarding each of the noted figures as location, time period, status (Tana, Amora),

specification of the name and all forms in which they appear in the sources (Roldán Vera & Schupp, 2006). Some entries also provide additional information including the appearance of the name in rabbinical literature, biographical information, and a list of the figure's masters, colleagues, disciples and family members.

The creation and analysis of social networks for historical figures has gained popularity in the fields of history and prosopography (Keats-Rohan, 2007); sociology (Roldán Vera & Schupp, 2006; Wetherell, 1998); and digital humanities (Rochat, 2015; Yamada, 2015). A study of a prosopographic social network is based on mapping group interactions onto graphs that represent individuals (nodes) and their relationships (edges). This execution enables one to inquire greater insight into the structure of social relationships and its impact on social group contexts such as activities, resolutions and behaviors (Novak et al., 2014). The social network facilitates the identification of the primary figures in the network in accordance to the number of compositions and frequency (Rochat, 2015; Keats-Rohan, 2007). It also enables the identification of groups on the social network, calculation of frequency and centrality; and extraction of additional network indices (Blondel et al., 2008). In the current study, we attempted to create a prosopographic network of the Sages of the Mishnah and Tosefta and to compare and contrast them.

## **Research Goals**

The study presents a new generic approach to identifying and characterizing editing styles of similar or parallel historical essays based on the construction of quantitative network profiles of the essays and statistical comparison between them. To this end, we developed a semi-automatic method for building a network of the characters that appear in each of the examined essays, we built network profiles of each essay and compared them by the quantitative metrics of these networks in order to identify significant differences between the editorial styles of the essays that cannot be identified simply from the structure or their textual content.

Our goal is to examine whether a quantitative evaluation of network profiles of historical texts can provide the information required to characterize their editing profile and reflect differences between their editing styles. As a case study, we applied the proposed methodology to two essays from the field of early rabbinic literature, the Mishnah and the Tosefta, which are very similar both in structure and in terms of the characters and topics they present.

## Methodology

This study is the first of its kind to utilize methods from natural language processing and methodic infrastructure of universal dependences (UD) – a framework used to clarify morphosyntactic elements of the human language that is generally employed to process text from spoken language – and implement them into the text of ancient Hebrew manuscripts.

In the first phase, we divided the corpuses into sentences, with each sentence represented by a word vector. In the second stage, we manually created a lexicon based on the UD scheme in order to extract information regarding the Sages from these word vectors. The words in the lexicon, which is entitled TanaimLex, express verbs, descriptions of Tanaim, locations and some of the names of the Tanaim. Every word in the lexicon is linked to its part of speech, which facilitates the understanding of the syntactical meaning of the word. Thereafter, we built rewrite rules as described in the UD scheme. Using these rules and the lexicon that we created, we successfully tagged keywords in the text using the parts of speech in the lexicon.

In the third phase, we sought to identify the names of the Sages. For this purpose, we activated functions to identify similarities and compare the names of the Sages in the tagged text to a file that encompasses the names of the Sages. The names of the Sages were identified, and each Tana received an identification number.

In the fourth phase, we extracted relationships between the Sages based on the type of connection – supplementary, supportive or citational. In this phase, we developed two algorithms, each operating using a distinct method:

The first algorithm was designed to locate associations based on lexical-syntactical patterns. Each pattern was tagged according to its part of speech as described above in Phase Two; and the algorithm extracted associations identified with the designated pattern from the text. Ultimately, this algorithm was less suitable for the manuscripts from the study corpuses, since they contain numerous writing errors; and we thus developed a second algorithm to identify associations based on sub-sequential categorization from the defined word vectors. A word vector containing all words of one Mishnah/halachah/law is called a complete sequence.

In order to divide the vector, we defined a predicate with all its dependencies which served as punctuation in the sentence. Any word sequence within a sentence that was suited to define a

predicate and its dependencies was divided into a sub-sequence and received a category code based on its predicate and dependencies. Every set of category codes from a pair of subsequences in a sentence named the association type (supplementary/ supportive/citational) that best suited it. Upon evaluating the findings, we revealed that the second algorithm, which identified and categorized sequences, offered improved results.

The algorithms were written in Python, and the visualization tools made use of Tableau, Excel and Gephi. In order to research prosopographic social networks of the Mishnah and Tosefta networks, we made use of common statistical network indices.

## **Primary Findings**

In the "Findings" chapter we supplied answers to the study questions by comparative analysis between the Mishnah and Tosefta. By comparing the distribution of appearances of the Tanaim in the text, we revealed that both the number of Sages and number of relationships in the Tosefta is higher than the number of Sages and relationships in the Mishnah. Moreover, nearly half of the Sages and relationships are common to both corpuses. Supporting relationships revealed very few shared relationships.

The analysis of prevalent Sages revealed that the prevalent sages appear more frequently in the Tosefta than in the Mishnah. It appears that Rabbi Yehuda bar Ilai is the dominant sage

in both corpuses, followed by Rabbi Meir Baal Hanes, Rabbi Shimon bar Yochai, Rabbi Yosi ben Halafta, Rabbi Eliezer ben Horkenus, Beit Hillel, Beit Shammai and Rabbi Akiva. Analyzing the relationships between prevalent Sages in both corpuses, we revealed that the most prominent supplementary relationship exists between the Beit Hillel and Beit Shammai. In supporting relationships, the only association found was between Rabbi Shimon bar Yochai and Rabbi Shimon ben Yehudah; and in citations, two relationships were found between Rabbi Eliezer ben Horkenus and Rabbi Yehoshua ben Chananya, and Rabbi Tarfon and Rabbi Akiva.

When presenting the Tanaim by era, the largest number of Sages in both corpuses was found in the period of Yavneh which was in the era of Bar Kochva's rebellion. It appears that the Tosefta encompasses more cross-generational relationships than the Mishnah. In addition, it is readily apparent that the percent and number of cross-era relationships in each corpus is low relative to the total number of relationships.

Analyzing the division of Tanaim and relationships based on a Power Law graph, we reveal that there is, indeed, a small number—up to fifteen—of very prevalent Sages/ relationships, while the others appear only sporadically in each corpus.

Comparing the Mishnah network to the Tosefta network, the dominant sages that appear from the analysis of prevalent Tanaim are prominent in both networks. Moreover, we revealed that the Mishnah contains concentrations of relationships as opposed to the relationships scattered across the Tosefta.

An analysis of ‘mediating Tanaim’—figures who transmitted halachic tradition to future generations—revealed that, in the Mishnah network, Rabbi Akiva received the highest mediating index. The communities of Beit Hillel and Beit Shammai, despite being an associated group, possess primarily internal associations and are less linked to other groups, which is why their mediating power is relatively low.

In the Tosefta network, Rabbi Yehuda bar Ilai occupies first place as mediator, with even greater power than Rabbi Akiva. Furthermore, it appears that the leading Tanaim in the Tosefta were also the leading Tanaim in the Mishnah, albeit unlike the Mishnah, the Tosefta encompasses many more Sages linked to other groups who possess significant mediating power.

When analyzing the division of relationships of the Tanaim in the profile of parallel tractates between the corpuses of the Mishnah and Tosefta, we revealed that several tractates feature a significantly higher number of relationships. It was further revealed that there is a high percentage of overlap in the relationships of Tanaim in specific tractates, whereas other tractates reveal no overlap at all. Comparing the percentage of overlapping Sages within parallel tractates in the Mishnah and Tosefta, we found tractates in which the number of sages was significantly higher. In contrast, there are several tractates in which the number of sages in the Mishnah is still greater than in the Tosefta. In the Orders of Kodshim and Taharot, it appears that the percentage of overlapping Sages is higher than the percentage of Sages in other orders.

Calculating the correlation between the appearances of Tanaim in both corpuses in the profile of parallel tractates, we revealed a strong positive relationship between the appearances of Tanaim in the Mishnah and the appearances of Tanaim in the Tosefta.

In addition, we found a single maximal clique in the Mishnah comprised of the dominant Tanaim including Rabbi Yehuda bar Ilai, Rabbi Eliezer ben Horkenus, Rabbi Shimon bar Yochai,

Rabbi Meir Baal Hanes, Rabbi Yosi ben Halafta and Rabbi Akiva. In the Tosefta, we found one maximal clique comprised of the same Tanaim noted above, with the exception of Rabbi Meir Baal Hanes.

Based on the entries of extractions and precisions, we found that the sequence identification algorithm provided more entries relative to the pattern identification algorithm. The sequence identification algorithm revealed higher percentages in supplementary relationships compared to a low percentage of citation relationships (in the Mishnah and Tosefta) and supporting relationships (in the Mishnah.)

### **Significance and Contribution of the Study**

This doctorate researched, for the first time ever, the two central manuscripts from the Tanaic era that contain similar structure and content: The Mishnah and Tosefta. These masterworks, which constitute the primary Rabbinical literature, represent the foundational development of Jewish law. In this study, we endeavored to evaluate their similarities and differences in terms of the representation of Sages and their opinions and the relationships between Sages and groups of Sages that they formed. It should be noted that the wide-ranging similarities identified in the two works, despite being compiled and edited by different parties, indicate factual historic information both from a social perspective and regarding connections between Sages.

The findings of this study are likely to facilitate the study of Talmudic literature and development of Jewish philosophy throughout the past two millennia, evaluate the impact of the editing of these works, and contribute to the study of Rabbinical literature from several perspectives:

First, the primary contribution of this study is the digitization, unification and analysis of data regarding the Tanaim from various sources of information. The collected data includes familial associations such as son, father, Rabbi, student, residence, generation, and data regarding associations between the Sages—supplementary, supporting and citational.

Second, this study reveals the potential of making raw and often inaccurate text from manuscripts accessible using a clear methodology. The methods presented enable one to process the manuscript beyond its syntactical level and to arrive at significant semantic definitions.

Another important benefit of this study is rooted in the increasing need to receive prompt answers to questions posed by students and researchers regarding the above manuscripts, such as: Who are the dominant Tanaim? Who are the Tanaim that connect between various groups and transmit Jewish tradition and halachah to future generations? Who were the intergenerational and cross-era mediators? Were there any cliques, and who were the figures who comprised them? Following the digitization and presentation of the manuscripts according to the presented methods, it is possible to resolve these questions using a simple query instead of conducting a manual search of all existing manuscripts.

In addition, the prosopographic database constructed enables the wide-ranging quantitative analysis of the networks of Tanaim formed by the data in the two works and their visual representation in a range of profiles and resolutions. It is now possible in a single comprehensive glance to review the texts with the goal of interpreting the phenomena and trends found in this literature. For example, in the distribution of the Tanaim in the profile of parallel orders and tractates in the Mishnah and Tosefta, we revealed that in the Orders of Kodshim and Taharot, the percentage of overlapping Sages was higher than the percentage of overlapping Sages in other orders. When we expanded the evaluation of this phenomenon to groups of Tanaim, we noted that Sages with the highest indices dwelled during the era of the Temple or learned firsthand of its procedures from eyewitnesses who had lived during that era. These findings reinforce theories and assumptions deriving from the study literature.

MMS Number: 9926892712805776