

The Sunbelt 2013 Data: Mapping the Field of Social Network Analysis

Jürgen Pfeffer

*Carnegie Mellon University,
Pittsburgh, Pennsylvania*

Betina Hollstein

*University of Bremen
Bremen, Germany*

John Skvoretz

*University of South Florida
Tampa, Florida*

Please send all correspondence to jpfeffer@cs.cmu.edu.

1. Overview

The Sunbelt Conference is the annual meeting of the International Network for Social Network Analysis (INSNA). In May 2013 (5/22–5/26), the 33rd Sunbelt Conference was held in Hamburg, organized by Betina Hollstein, Sonja Drobnič, and Michael Schnegg. The keynote address (Simmel Award) was delivered by John F. Padgett (“Networks and History”). The Freeman Award lecture was presented by David Schaefer (“Distinguishing Pattern from Process: Equifinality and Network Selection”). 1,050 researchers from 50 countries participated at the XXXIII Sunbelt Conference to watch 656 paper presentations in 127 sessions (see network visualization in Figure 1) and 93 posters – overall 749 research products from 1,351 different authors.

Looking at the number of participants as well as the number of presentations, Sunbelt XXXIII was the largest conference in INSNA history so far. Participants represent a diverse array of scientific fields and cover a broad variety of topics. Consequently, analyzing the content of the conference can provide us with deep insights into the state of the art of research in the field of social network analysis.

This article describes the dataset that represents the printed conference program (including abstracts and keywords). We also describe the process of creating the program from the abstract submission system to the final placement of paper presentations to sessions.

2. Data Collection

For collecting the submissions of the paper and poster presentations, we turned to a market leader from Germany in the conference administration field, “pharma service” and used their online tool “Full Service Abstract

management”.

The submission system closed (after being extended for one week) on January 7, 2013. 848 abstracts have been submitted by researchers from around the world. This number includes a small number of abstracts that have been submitted slightly too late due to technical problems. During the submission process a set of information (authors, institutions, abstract, keywords, etc.) has been collected. These variables will be described in Section 3. We start out with the process of assigning paper presentations to sessions (time slots), as this is the main challenge of creating a Sunbelt program.

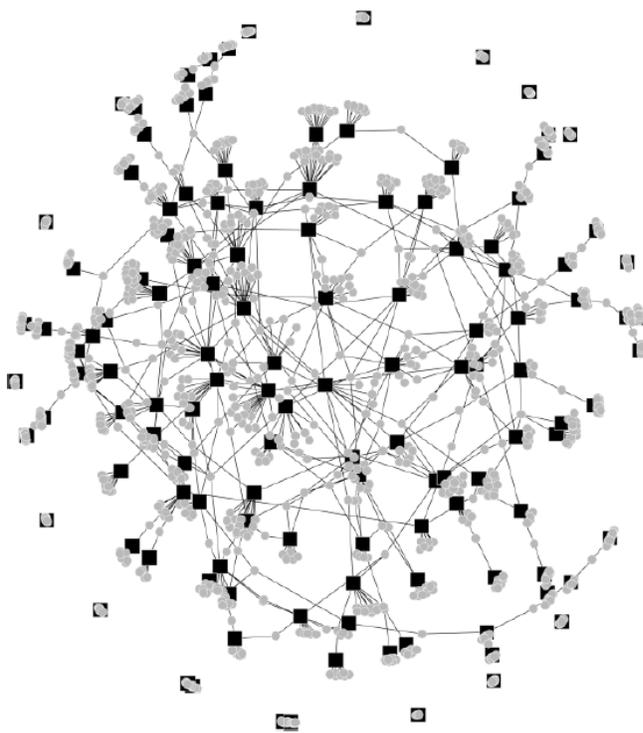


Figure 1: Sunbelt 2013 conference - sessions and people

2.1. Assigning Paper Presentations to Sessions

During the submission process, every person submitting an abstract was asked to select a preferred presentation form, namely either a 20 minute paper presentation or poster presentation; a third option was “no preferences”. For the majority of submissions, the paper or poster selection is identical with the final presentation form; some who indicated the poster option in the special note section of the submission form have been altered accordingly by the local organizers in order to get hold of the large number of submissions. Most submissions indicating “no preference” were assigned to the poster session for the same reason. In this article, we primarily discuss paper presentations. Although, poster presentations are also included in the data (see next section).

For a paper presentation, submitters were asked to suggest a possible session topic¹ for the talk. A pre-defined selection of 72 session topics was compiled for this purpose based on sessions held at the last three Sunbelt conferences. However, it was also possible to create a new suggested session topic.

32 of the pre-defined session topics were organized, i.e., researchers had agreed on assisting in the session assignment process as well as on chairing the respective sessions at the conference. Most of the organized session topics had separate call for papers via the SOCNET email list and the conference website. 412 (=48.6%) of the submissions were covered by organized session topics.

It is important to notice that these suggested session topics are not always identical with the final session assignment. During the program creation process, some paper presentations were moved from organized to other session topics. Finally, session topics with lots of submissions were split to multiple sessions, and session topics with few submissions were joined. Beside the goal of putting similar abstracts in one session, there is another major constraint for the session assignment procedure. In general, the number of talks per session is fixed to five or six, except for a small number of sessions (normally the first session on the first day and the last session on the last day of the conference) less paper presentations are possible.

2.2. Rejects and Withdrawals

From 848 initial submissions 99 (11.7%) have been removed in the time period from January 7 to April 27, 2013. There were three reasons for a submission not getting into the final program:

- Rejected by the Sunbelt 2013 organizers if the submission did not comply with the minimum requirements² and the authors did not provide a recasted version of the abstract after being asked to do so. (9)
- Rejected because of missing conference registration (28)
- Withdrawn by the authors (62)

This Sunbelt 2013 dataset reflects the state of the printed program (deadline April 27, 2013) including 656 paper presentations in 127 sessions and 93 posters. Later withdrawals (that have been removed from the interactive online program³) remain in this dataset.

3. Data Files and Formats

The Sunbelt 2013 data consists of several tables. The tables are stored in separate sheets of a single Excel file (“SUNBELT 2013 Data.xlsx”). These tables were used to create the printed and the interactive online program. There are two key variables that connect the tables and both are essential to the program:

ID: A unique identifier for every submission. No overlap between paper and poster presentations. IDs connect authors, titles, affiliations, keywords, and abstracts in different data tables.

Session Code: A code that connects a presentation with a specific session. These codes are constructed from four parts, the weekday, morning or afternoon, first or second session (see sheet “TimesSlots”), and the room ID.

4. Data Details

In the following, we describe the data tables of the Sunbelt 2013 data. In general, data columns are in original form as provided by the researchers when submitting their abstracts. Post-processed data are reported explicitly. Please notice that for both paper and poster presentations

¹We use the term “session topic” to take into account that some of which resulted in multiple sessions.

²Sunbelt XXXIII Conference Guidelines & FAQs, Abstract Submission FAQ, page 9: “The abstract must describe some work that is about NETWORKS, most likely, social networks...”

³www.insna.org/program2013

a one-presentation-per-person rule were carried out and that the first author is not necessarily the presenter.

4.1 Paper Presentations

This table includes all 656 paper presentations. Every data line represents one submission. Here are the columns of the table:

<i>ID</i>	Submission identifier
<i>Abstract Title</i>	Title of the paper presentation
<i>Author(s)</i>	Author(s) of the submission. The presenter is underlined; superscript numbers connect people to institutions in case there is more than one institution involved.
<i>Institution(s)</i>	Institution(s) of the author(s); superscript numbers connect to author(s).
<i>Country</i>	Country of the person that submitted the abstract. The person doing the submission is not necessarily the presenter or the first author.
<i>Session Title</i>	Title of the session in which the paper was presented. This is the assigned session (see section 2.2.) not the session topic suggested by the author(s).
<i>Session Code</i>	Day/Time slot and room ID describing when and where the paper was presented.
<i>Talk Nr</i>	One session consists of multiple talks (normally five or six). This number indicates the position of the paper presentation within the session.

4.2. Poster Presentations

This table includes all 93 poster presentations. Every data line represents one submission. The columns in this table are identical to those described in the previous section (paper presentations) except that there are no sessions assigned to posters as all posters have been presented in a single poster session.

4.3. Keywords

This table shows the selection of keywords by the people submitting the abstracts. Every keyword/submission link is a single line to make it easier to import this table into a network tool. The selection of keywords was restricted in the submission system by two factors. First, the keywords were pre-defined and no new keywords could be entered. A list of 100 keywords was compiled from keywords used for the last three Sunbelt conferences. Note that there is also a keyword “others”. Second, the number of possible keywords to select ranges from 1 to 5. 2,813

keywords have been selected (avg. 3.76) for all 749 paper and poster presentations. Every single keyword was used at least two times. The top used keywords are Social Capital, Egocentric Networks, and Inter-organizational Networks. The columns of the keywords table are defined as follows.

<i>ID</i>	Submission identifier
<i>Type</i>	Paper or poster presentation
<i>Keyword</i>	Keyword selected from a pre-defined list of keywords

4.4. Institutions

For further post-processing of the authors and institutions outside of Excel, the superscript affiliation numbers can turn to an intricate problem. This table should ease this data manipulation pain. Still, in order to get your institutional collaboration network, some cleaning will be necessary as these are text data that were typed separately for every submission. The columns of this table are copies of previously described tables.

4.5. Index

This table was created for the index of the printed program. This table (excluding the posters) was also the data source for the Sunbelt 2013 conference poster of sessions and people (Figure 1).

Every line connects one person with a specific session represented by the Session Code. The names in this table have been cleaned so that different writings of one person result in a single name, e.g. “Pattison, Philippa”, “Pattison, Philippa E.”, and “Pattison, Pip” were all converted to “Pattison, Philippa E. (Pip)”. This cleaning process was just done for this table and did not change the writing of the authors in other tables. Note that all posters have been coded with “Poster”.

<i>Name</i>	Author of a submission (cleaned).
<i>Session Code</i>	Day, time slot, and room ID of the session in which the author gives a paper presentation

4.6. Abstracts

749 abstracts are part of the printed program. Every abstract can be found in a single line in this table. 157,000 words of these abstracts added up to the 300 pages “Abstract Program” that can be found online at the conference website. Be aware that some researchers like to prepare their abstracts in MS Word or other text processing tool and then copy/paste the abstract—including formatting

and special symbols—to the abstract submission system. If you are planning to analyze the abstracts, be aware that most of the original formatting (including paragraphs) has been deleted in the data handling process; sometimes by leaving enumeration symbols in the middle of the text, e.g., see submission 457. Also pay attention to the fact that older Excel versions crop the abstracts due to a limitation to the number of characters by cell.

<i>ID</i>	Submission identifier
<i>Type</i>	Paper or poster presentation
<i>Abstract</i>	Abstract of the submission

4.7. Sessions

The session table describes when and where a session was held and if that session had organizers (see section 2.2). In case of an organized session, the organizers served as session chairs, otherwise the last presenter of the session was assigned as session chair. The maximum number of talks per session is six. Multiple sessions that are the result of a single session topic (see section 2.2) can have enumerations (e.g., Social Capital 1, Social Capital 2) or title extensions (e.g., Words and Networks: Health and Culture, Words and Networks: Politics and Crises, ...).

<i>Session Code</i>	Day, time slot, and room ID of the session
<i>Session Title</i>	Title of the session
<i>Day</i>	Day of the conference (Wednesday=1, Sunday=5)
<i>Weekday</i>	Wednesday, Thursday,...
<i>Slot</i>	Time slot of the session (see table <i>TimeSlot</i>)
<i>Room</i>	The ID of the room where the session took place.
<i>Organizer(s)</i>	Name(s) of the organizer(s) in case of organized sessions

4.8. Times Slots & Paper Presentation Times

In order to create the program, two more tables were necessary. The table *TimeSlots* defines the time Slot entries of the tables described above and the table *PresentationTime* maps a given time Slot and a Talk Nr combination to the exact starting time of the paper presentation.

5. Additional Materials

A high-quality printable PDF version of the Sunbelt 2013 conference poster of sessions and people as well as the

related network file can be found here: www.pfeffer.at/sunbelt2013

Acknowledgements

Several people have been engaged in creating this dataset. Dagmar Zanker compiled the keyword list as well as the list of suggested sessions for the registration system. Both lists are based on the work done by Laura Koehly, Rebecca Davis, and Tom Valente, the organizer of Sunbelt XXXII (2012) in Redondo Beach, California. Julie Hewett (INSNA) managed the registration system of the conference. Sonja Drobnič and Michael Schnegg reviewed the abstracts together with Betina Hollstein. 61 network scientists organized sessions and helped with assigning talks to sessions. Tom Töpfer helped clean the names for the index. Finally, thanks to the many people that actually wrote and submitted the 749 abstracts and presented their work at Sunbelt 2013 in Hamburg.