**Oz Shapira & Amit Tiroshi**

# Introduction

What if someone will tell us that he love pizza ,James bond movies and he also love to beach jogging and we can replay to him that he probably love to bake, hike and flower interweaving. This research will try to constant a valid recommender system based on graph analysis with the ability to answer on those kinds of questions in high absoluteness.

Recommender systems are part of user modeling systems with Recommendation abilities, those system “knows” according to past users knowledge how to make recommendation according to his flavored .

Recommender system applies knowledge discovery techniques to the problem of making product recommendations during a live customer interaction. These systems are achieving widespread success in E-commerce nowadays, especially with the advent of the Internet .there is many type of recommendation systems and in each one of them have different approach for recommitting (vs. Pandora , Google search ,YouTube ,amazon… ) There is two type of recommendation system:

**Content-Based**, - where recommendations are based on semantic properties (preferences) of the items (users).

**Collaborative-Based** - where the recommendations are based on previous ratings of (similar) users to (similar) items, with the assumption that users who agreed in the past on item ratingsare likely to agree again in the future.

<search for more if more is need >

but most of those system are need to obtain user profile for their recommendation ,some of those system are an application of a particular type of Knowledge Discovery in Databases (KDD) Fayyad et al. [1996] technique how investigate the useable database for discovery the “best option” for the recommendation.

// delete or remove to background –

At the World Wide Web many sites contain user’s profile which reflects user characters.

In some of the site this data have been add by the user itself while in different sites data is obtain independently , for example in Facebook the user insert is interests while in Pandora (Pandora its music recommended radio website system with broadcast music according to user flavor) try to find user flavor according his user music analysis.

In most cases user model systems are trying to obtain user knowledge, learn their styles and individual traits, research on this mattes have been concluded by p.brusilovsky & millan [2007]. the more the system is open, generic and shareable its increase the possibility for more accurate user profile, the main problem in user modeling is attempting to modeling new user when we have very little data on him, this research outcome can assist on this problem by identify individual traits using collection of many user traits (at the first stage ) and find the connection between those traits and applied ontology for new users .

A type of systems that analyzing the user data reference and recommend to user on new user preference called recommendation systems,

//end

In most knowing recommender systems today are specializing on one domain (for example Pandora as music recommender, Netflix offers predictions of movies) and simply called domain recommendations system , the ability to focused on single domain give to those systems the focused they need for correct proposals. Usually those system are using single dimensional database or many databases with common fields therefore most of those systems have been successfully develop algorithm for collecting and analyzing their data.

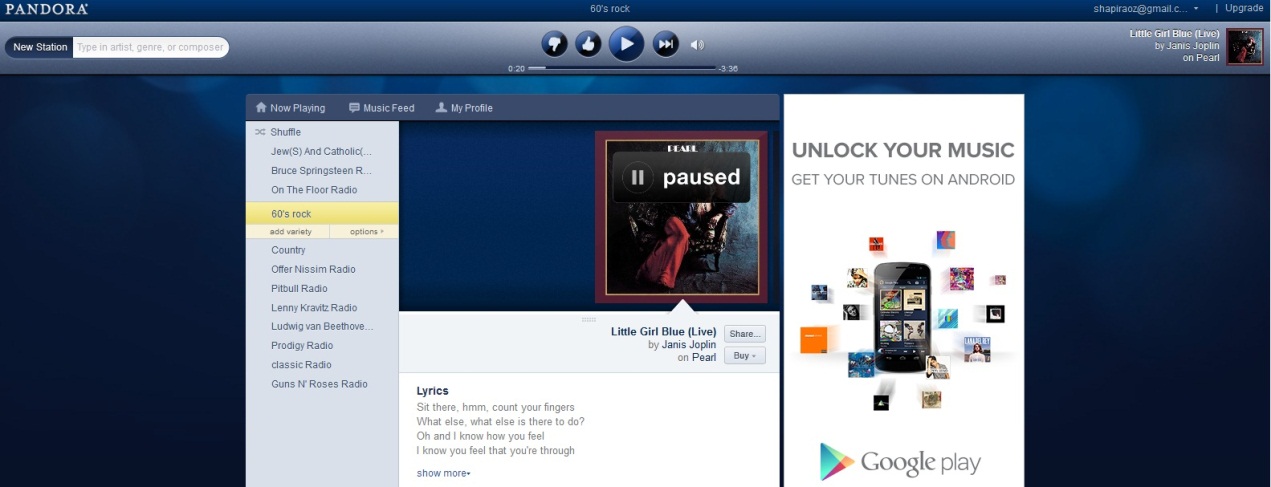


Figure 1: Pandora as recommender system

In our research we will try to construct cross domain recombination system with the abilities to applies recommitting to cross domain , this task Is very complex unlike single domain the varied of domains in unlimited , today each human have different aspect of interest and came from different culture which create different type of characteristic. Therefore for obtains user characteristic the process should be open to differential in users traits. in addition to that domain can by misunderstand form user context for example if user love animal he can create at is site profile “animals are great“ ,“ my loving animals” or just “animals” in all this cases we can understand that our user love animals , but what if will write “the cutest animals ” or “wild animals” – it’s have different meaning . Cross domain system should be analyze the differential between domains and its subdomains.

<need collecting problem…>

the effort of collecting data have been mention before by [S D Rhodes](http://jech.bmj.com/search?author1=S+D+Rhodes&sortspec=date&submit=Submit)[1](http://jech.bmj.com/content/57/1/68.short#aff-1), [D A Bowie](http://jech.bmj.com/search?author1=D+A+Bowie&sortspec=date&submit=Submit)[2](http://jech.bmj.com/content/57/1/68.short#aff-2), [K C Hergenrather](http://jech.bmj.com/search?author1=K+C+Hergenrather&sortspec=date&submit=Submit)[3](http://jech.bmj.com/content/57/1/68.short#aff-3) [2003] in their research they concluded that using the web as empiric tools for behavioral science research will increase the tested population from local to global distribution . Using the web as resource for data mining is not new and used in many researches, in our effort we will try to found normally distribution population for secure or data misleading, for this effort we can find many type of populations at the socials networks.

Social networks have been with us since 1997 (the first one was sixDegreee.com) , social networks site (SNS) have successfully change worldwide communication they gave personal user the ability to reach any user in the world , SNS site attracted millions of users, many of whom have integrated these sites into their daily practices. As of this writing, there are hundreds of SNSs, with various technological affordances, supporting a wide range of interests and practices (for example Facebook) those abilities cause to SNS the basic ability to connect between separate type of population using SNS users ,danah m boyd and nicole elision [2007] rise the fact the SNS can provide rich sources of naturalistic behavioral data. Profile and linkage data from SNSs can be gathered either through the use of automated collection techniques or through datasets provided directly from the company, enabling network analysis researchers to explore large-scale patterns of friending, usage, and other visible indicators ([Hogan, in press](http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full#b45)), and continuing an analysis trend that started with examinations of blogs and other websites .

SNS are basically contain social circle when each one of those circle can related to different aspect for example in Linkin (Linkin is professional SNS that specialize work relation) most likely to normal user will have work circle (he will have connection to people in is work) but he also can be at different circle for example is friend from school and is army service. At same way it’s can happen in Facebook user have friend from different circle : school, university, work place, neighborhood, music he love , food etc… in Google+ they even coded this feature as you can create or join to circle – when each circle have the common topic.

# Background and Related works

## Seeking for normally distribution source

Since we interested of creating affective database with reflect all type of population the collecting must be breath spreading and must have direct links between some of individual , it’s also must contain many domains with some kind of link between them ,the source must also be dynamic, scalability and can be change .

## Social Network as a source

We found that the social networks (SN) can be effective sources for establish database, the main key in social networks is to shard the individual to the common population. That value can attribute to our goal, each social network are depend user data, in general at SN user are upload their data to the SN, they are deiced how to shard (to the common population, to your link friend only or save as private to your used ) also can benefit the free cataloging – when users are update their data they choose how to call and catalog it form this act we enumerate the SN as follow : when each user is update some data, we can refer it as new object (new field) in the database the name of that field is the name that the user is catalog it , for example : Let us suppose that some random user at Facebook with cooking hobby , some day he decided to bake special birthday cake for his person in his family, when he finish making cake he decided to photograph the cake and upload the picture and the recipe to Facebook , he also label “ homemade Birthday cake for <person >” from that act we can examine the this picture as object , we can understand that some person have birthday , we can also understand that this user love to bake ,and it’s person love cakes, when different users are press like for that cake we can also assume different conclusion for example :we can assume that users how press like or send message to the cake are probably people that love cakes or at least love the shape of that cake , people that will press like on the recipe most likely people that love to bake to. This case is one many cases that happen hundred or maybe even thousand time at each day, the conclusion that we mention before are easy for human to understand, but we need to save those conclusion and save the result in database.

Additional aspect that tribute for our effort it’s SN’s are daily even hourly updated, danah m boyd and nicole elision already mention that SNS combine form millions of users, we can construct automated systems that can give a “big picture”- large scale analysis that can daily update and synchronize with SNS , also since SNS are basically web system we can access to their data from anywhere (it depend on site access – we will enlargement at section 2.4) since we base on data that exist in the computerize system – we can create system that can used the SNS interface for extracting relevant data. Using automated system like we describe will allow collecting anytime anywhere data.

Fehmi Ben Abdesslem, Iain Parris, and Tristan Henderson (2011) research and concluded the use of SN for collecting data, they speared the collecting to two sections – collect user social behavior in SN and collect user characteristics. from their aspect when we used SN we collect not only is preference and characteristics we can actually build social profile from the user data, in their research they collect data from SN , but in their twitch they not collect the data only from SN the also use mobile devices and for crossbreed their collected data with user location , by this way they success to “understand ” user act’s ,behavior and desire . they successfully collect data , but the main benefit that we can concluded is the importance of random sampling , Fehmi ,lain and tristan make excrements with that interacts with user and used Facebook as the SN , for creating random sampling they select users for his friend , their application assist with “all friend” user and extract new user – with this approach they achieve random sampling and keep their subject normally distributed.

## Connection Analysis

At the same way Fehmi ,lain and Tristan we also have to make sure that our subjects will be normally distributed and **not** belong to some type or circle in the SN but unlike fehmi we don’t want to collect random data from user , at the end we want to understand the nature of the connection between the user and their preference ,therefore we will have to collect that data in some way that allow us to understand the native of the connection between each preference and characteristics for building characteristics connection graph.

In the birthday cake example that we mention in section <> user can connect each other – this in one of the strongest abilities exist in SN , the fact that users can add link to each other (via friend in Facebook) is create a grid of users with connection to each other – we can look at this grid as graph and use it for creating characteristics connection graph example , in this research our main purples is to collect the data , but also to arrange in in graph , the characteristics graph have to build dynamically and update form each user at the SN , therefore the process of collecting the data is automatically we be depended on the relation with SN users .

## Related works

### Generic Semantic-based Framework

Ignacio Fernández-Tobías , Marius Kaminskas , Iván Cantador and Francesco Ricci public article (A Generic Semantic-based Framework for Cross-domain Recommendation 2011) try to create automated system that will recommend to user preference by to different domain in their approach the used graph for mapping the connection between the domain and analyzing the nodes relation in graph , this approach adopt the *Content-based recommendations* mention on paper by Gediminas Adomavicius1 and Alexander Tuzhilin (Towards the Next Generation of Recommender Systems:A Survey of the State-of-the-Art and Possible Extensions -2005) used past user data for create the recommendation ,in their system they used DBpedia as the database source. BBpedia is graph base database that obtain is values from Wikipedia , the main problem with that experience is DBpedia is not updated Daily , in fact it have version that come out once in some time ,since they can be depend on that database they created description framework built upon semantic networks , the problem with attitude it’s for obtaining data you to maintain and upgrade your frameworks.

### Link Data cloud

Figure 2: Fragment of the Linked Data cloud

Within the Semantic Web initiative, the Linked Data5 project aims at publishing structured datasets – usually described by standard metadata models such as RDF6 – on the Web, and setting (RDF) links between data items – usually called semantic entities – from different data sources. The adoption of Linked Data has thus led to the extension of the Web with a global data space connecting data from diverse domains such as people, companies, books, films, television, music, statistical and scientific data, and reviews [5]. This enables new types of applications. For instance, there are search engines that crawl Linked Data by following the links between data sources, and provide expressive query capabilities (see e.g. SPARQL7 RDF query language) over aggregated data, similar to how a relational database is queried today.

### Collecting data by SNS

Fehmi Ben Abdesslem, Iain Parris, and Tristan Henderson (2011) used the social networks and mobile devices for collecting user characteristic and understand in behavior, in their research they used Facebook API for obtaining users data form Facebook while the activate questionnaire at user mobile devices, they also crossbreed their data with user location (they use GPS in the mobile device) for “understanding” here user and collect effective data and avoiding bad filtering.

In this paper we will try to obtain database by using the SNS system the biggest question is since SNS are combine for many type of the population when each item is basically related to some circle how we sample subject in random way without taking our samples from the same circle in this paper will try to answer this question.

## Summery

<Need to complete>

# Research Goals and Questions

The goal of the proposed research is prove and to answer on the following question:

* How to create recommendation system based on graph when data source is curator Social network
* How we can use the SN for collection user’s traits and interested? - In this question will explored way to collect form common population his hobbies and characteristic in automatic large scale way.
* Does mathematics Graph can used to map human traits and interested?
* Can we used mathematics Graphs for predict human traits and interested? – in this question we evaluated the Traits graph and check for
* What kind of information we can extract from this kind of graph and how to obtain this graph for valid recommendation algorithm?

# System and Methods

We will construct a research tool called **TraitsFinder** that will allow us to collect user’s data and extract from this data mathematical graph that will represent the connection between traits, this tool (Traits Finder) will crawled SN (pinterest) and collect user’s traits in our servers .

At the second stage is graph building , in this stage we will add to our tool graph building abilities with basic analysis.

<Need to Add>

## Methodology

## 4.2 Evaluation

# 5. Timetable

Each of the phases discusses the primary focus of each time period:

Phase I - Literature Survey and Focus.

Phase II - establish first stage of TraitsFinder – the web crawler with simple graph builder

Phase III - collecting users traits (exit criteria: at least 100,000 users)

Phase IV - graph investigation .

Phase V - improve TraitFinder: add automatic graph analysis GUI and RI

Phase VI - Graph analysis, evaluation and algorithm establish

Phase VII - Writing Thesis .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Phase | summer 2012 | Fall  2013 | spring  2013 | summer  2013 | fall  2014 |
| I |  |  |  |  |  |
| II |  |  |  |  |  |
| III |  |  |  |  |  |
| IV |  |  |  |  |  |
| V |  |  |  |  |  |
| VI |  |  |  |  |  |
| VII |  |  |  |  |  |

# 6 Research Contributions

# 7 References

1. Fehmi Ben Abdesslem, Iain Parris, and Tristan Henderson (2011)- “Reliable Online Social Network Data Collection”
2. Ignacio Fernández-Tobías , Marius Kaminskas , Iván Cantador and Francesco Ricci (2011) -, “ A Generic Semantic-based Framework for Cross-domain ecommendation”
   * 1. danah m. boyd1,•, Nicole B. Ellison (2007) - “Social Network Sites: Definition, History, and Scholarship “
3. peter brusilovsky and eva millan (2007) – “The Adaptive web : user models for adaptive hypermedia and adaptive Educational systems”
4. Adomavicius1 and Alexander Tuzhilin (2005) - “Towards the Next Generation of Recommender Systems:A Survey of the State-of-the-Art and Possible Extensions”
5. S D Rhodes1, D A Bowie2, K C Hergenrather3 (2003) –“ Collecting behavioural data using the World Wide Web: considerations for researchers”
6. Hogan, B. (in press). “Analyzing social networks via the Internet. In N.Fielding, R.Lee, & G.Blank (Eds.), Sage Handbook of Online
7. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl (2000) “Application of Dimensionality Reduction in Recommender System – A Case Study”
8. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.,and Uthurusamy, R., Eds. (1996) - “Advances in Knowledge Discovery and Data Mining”
9. Rubi Boim (Under the supervision of Prof. Tova Milo) Tel-Aviv University “Methods for Boosting Recommender Systems”