Project Proposal
Project Name J.A.R.V.I.Z

Amish Goyal, Vinay Kola, Vinodh Krishnan, Vivek Nabhi

March 9, 2014

1   Problem Statement

Movies from Marvel and DC comics have proven highly successful in the last decade. It is easy to forget that these "comic-verses" actually have 70 years of history behind them. We plan to give fans of the movies a starting point for their comic reading journey, by helping them wade through the vast lists of characters and recommend the best story arcs for every comic book series.

2   Heilmeier's Catechism

1. What are you trying to do? Articulate your objectives using absolutely no jargon. What is the problem? Why is it hard?
Our project aims to enable people with minimal or no knowledge regarding the comic book universe navigate and appreciate it through interactive visualizations and relevant recommendations. This would help them begin their comic book reading journey without being overwhelmed by the 70 years of history and the dizzying list of characters. Most first time users are discouraged from reading these books simply because they don't know where to start. Developing an interface that enables users to navigate through this large dataset is one of the prime challenges of this project.

2. How is it done today, and what are the limits of current practice?
Comic book recommendations are done through "Recommended Reading" lists[14], like but these recommendations are black boxes. They do not give the user any context in terms of the rest of the comic book universe. There are no systems that recommend other characters based on user preferences. There have been visualizations of the Marvel Universe "social network", but the sheer scale of it has rendered these visualizations relatively pointless to the end user.
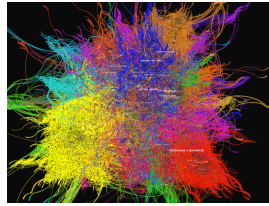
Figure 1: Graph illustrating the Marvel Universe network generated by [15]

3. What's new in your approach and why do you think it will be successful?

We endeavour to view the graph at various levels of granularity, by applying filters a la Graph Search for Facebook. By reducing the scale of the graph with the user-applied filters, it becomes more useful to the user. We also provide a list of recommended characters and an optimal order of story arcs that give the best perspective on their evolution.

4. Who cares?

The popularity of Marvel movies has brought a lot of new fans to the Marvel universe. However, it is difficult for them to comprehend the complexity of the comic book universe. Most fans want to start reading comics, but there are not many systems currently that recommend the optimal storyline for a user to follow. Also, the user might want recommendations for other characters similar to the ones they follow, which most current systems don't provide.(the social network paper gives some examples)

5. If you're successful, what difference will it make? What impact will success have? How will it be measured?

If successful, it will make the comic book experience for both first time and seasoned users very enjoyable. Our system provides an interactive experience for the user to drill down to a particular list of characters based on their interests, and then find the list of books that follow the main story arc and provide the sequence in which he/she should read them. There is also a possibility of monetizing this project by linking our recommendations to the corresponding comics on Comixology.

6. What are the risks and the payoffs?

One of the major challenges would be to handle the performance concerns that come with such a huge database. Additionally, employing content-based recommendation may not always yield the best results. We would like to tie in collaborative filtering to our recommendation system ideally, but this is outside the scope of this project.

7. How much will it cost?

All the APIs we plan to utilize for our project are available online, free of cost. As we plan to make our visualization web-facing, we may incur some costs for Amazon EC2 hosting.

8. How long will it take?

We estimate that it would take around 6 weeks to complete the project. A detailed breakdown is given in a later section.

9. What are the midterm and final "exams" to check for success? How will progress be measured?

By midterm, we plan to implement our recommendation system, although it may require some improvements. Regarding visualization, we plan to have a basic graph-based visualization prototype of all the characters implemented, so that we can evaluate any possible performance concerns. The final product should be a public website showcasing an interactive version of the visualizations described above.

3 Related Work

Marvel and DC Comics[4, 5] have been the two predominant publishing companies for decades now, and a side-effect of that is that their respective universes have accumulated a large swath of characters and comic book issues.

Alberich et al. [3] compare a social network constructed based on the Marvel universe with that of a real life network and find that they have many similar characteristics. Easley and Kleinberg[6] talk about different social network theories such as Structural Balance Theory that are applicable to real life networks. Since our graph has been shown to be similar, we can use such theories to make predictions even on the Marvel social network.

Eifrem[11] depicts the benefits of graph databases over traditional RDBMS in terms of size, connectivity, semi-structure and architecture. We find that since we are modelling a large graph, such a graph database would be the right choice. In fact, Marvel is building their own graph database solution for this problem. [7]

Koren et al.[10] discuss the matrix factorization approach as an improvement over the conventional nearest neighbour method by running experiments on the Netflix movie dataset. A similar approach might benefit in the comic book recommendation task. It would be important to carefully choose the features to represent these comic books.

Jomaa et al's[8] work is most similar to ours in terms of objective. It aims to recommend comic books based on both collaborative filtering and content-based features. We can incorporate the content-based features they used into our feature set. In future iterations of our product we could employ the user-relevant features like emotional and sensory features that are mentioned in the paper. In addition, we highlight the process of recommendation rather than treating it as a black box.

Lops et al.[9] highlight the state of the art techniques used for content based recommender systems. Some of the approaches that can be useful for our project include keyword based recommendations, using articles for comic-comic similarity etc. A potential source of problem could be the large number of bag of words

features that would be generated from texts related to comics. The dimensionality of the feature space would have to be controlled to get better results. Pang et al[12] performed sentiment analysis on movie reviews and summaries. In our recommendation system, we envision using the sentiment of the plot of the comic book as one of our features.

Herman et al[1] survey different types of layouts for graph visualization. It gives us a good idea about which graph visualizations would be suitable for different tasks, and which one out of those are relevant to us. Hyperbolic layout seems to be the most relevant for visualizing large datasets on a graph.

Shani and Gunawardana[2] extensively cover evaluation metrics for Recommendation systems. They cover different measures and techniques to conduct evaluation on recommendation systems including user studies and mathematical error predictions like RMSE and MAE. It is useful because it gives us metrics to evaluate our recommendation system with, which forms an integral part of our project. Schroder et al.[13] talk about improved evaluation measures for recommendation systems which correct the biases of the F-measure. It also gives guidelines for choosing between different metrics that will help us greatly in deciding the metric we want to evaluate our system with.
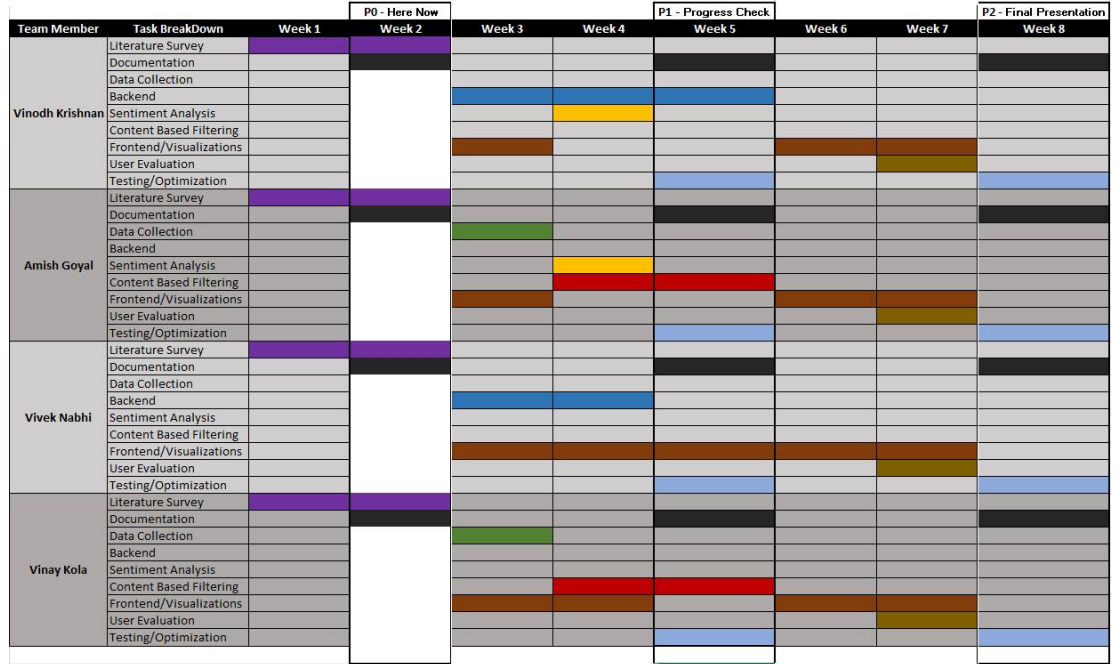
4   Timeline

Figure 2: GANTT Chart showing the work that will be done by each team

Papers read:
Amish Goyal- [3], [9], [10]
Vinay Kola - [1], [6], [12]
Vinodh Krishnan - [8], [9], [11]
Vivek Nabhi - [2], [4], [5], [13]

References

[1] Herman, Ivan, Guy Melançon, and M. Scott Marshall. "Graph visualization and navigation in information visualization: A survey." Visualization and Computer Graphics, IEEE Transactions on 6.1 (2000): 24-43.

[2] Shani, Guy, and Asela Gunawardana. "Evaluating recommendation systems." Recommender systems handbook. Springer US, 2011. 257-297. APA

[3] Alberich, Ricardo, Joe Miro-Julia, and Francesc Rosselló. "Marvel Universe looks almost like a real social network." arXiv preprint cond-mat/0202174 (2002). APA

[4] L. Daniels. Marvel: Five Decades of the World's Greatest Comics. Abrams, New York, 1991.

[5] M.R. Lavin. A librarian's guide to Marvel Comics. Serials Review, 24(2):47–64, 1998.

[6] Easley, David, and Jon Kleinberg. "Networks, crowds, and markets." Chapter 05. Cambridge Univ Press 6.1 (2010): 6-1.

[7] http://www.fastcolabs.com/3022482/marvels-new-uberframework-graphs-every-character-in-the-universe#3

[8] Jomaa, Inès, et al. "PREFERENCE BASED SYSTEM: AN ASSISTANCE FOR CHOOSING A COMIC." PREFERENCE BASED SYSTEM: AN ASSISTANCE FOR CHOOSING A COMIC (2011).

[9] Lops, Pasquale, Marco de Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." Recommender Systems Handbook. Springer US, 2011. 73-105.

[10] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 42.8 (2009): 30-37.

[11] Eifrem, Emil. "Neo4j—the benefits of graph databases." no: sql (east) (2009).

[12] Pang, Bo, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004.

[13] Schröder, Gunnar, Maik Thiele, and Wolfgang Lehner. "Setting Goals and Choosing Metrics for Recommender System Evaluations." Proceedings of the Second Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI 2). 2011.

[14] http://dc.wikia.com/wiki/Batman_Recommended_Reading

[15] http://exposedata.com/marvel/