

# Social Networking Study on DeviantArt

CmpE 491 Project

By: Ferhat Elmas

Student Id: 2006101102

Advisor: Haluk Bingöl

*19.01.2011*

*Version 6.0*

## VERSION HISTORY

Version	Date	Explanation
1.0	22.11.2010	Research definition and motivation
2.0	29.11.2010	Data fetch
3.0	06.12.2010	Addition of analysis algorithm
4.0	13.12.2010	Variances of algorithms and results
5.0	22.12.2010	Artificial Network
6.0	05.01.2011	Addition of sub life cycle algorithm and further research

## ABSTRACT

This project describes an explorative research into deviantART, a site that is created for sharing user generated artworks. The website offers various web-based services to its members enabling and enforcing a strong social interaction. With its collection of around 100 millions works, DeviantART (dA) is the biggest art market of the world.

In that sense, DA draw the way arts are generated, shared and enjoyed. The study is interested in enjoy function and focuses on the impact of what our idols enjoy. Idols are popular deviants who attract others with favorite feature. Favorite relation on dA is analyzed whether it can force deviants to choose deviations. Finding the most influential deviants on dA is the purpose of this study.

An intelligent crawler is written to collect the data. Algorithms that give deviants fixed and weighted coefficients are applied and for verification an artificial network is created. Algorithms give uniform effectiveness for deviants because of smoothing and calculations on whole life cycle of the deviations. Dividing time span increases the power to find the most effective user. However, this algorithms must be improved and found deviants should be followed in the community according to provisions.

# TABLE OF CONTENTS

VERSION HISTORY . . . . .	i
ABSTRACT . . . . .	ii
1. Introduction . . . . .	1
2. Difference of Current Project . . . . .	3
2.1. Why is Favorite - Watcher Relation important? . . . . .	3
2.2. How can Favorite - Watcher Relation shape the network? . . . . .	3
3. Data Fetch . . . . .	4
3.1. Wayback Machine . . . . .	4
3.2. Periodic Snapshots . . . . .	4
3.3. Intelligent Crawler . . . . .	5
4. Effect Coefficient - Algorithms . . . . .	6
5. Results of Applied Algorithms . . . . .	8
5.1. Improvements . . . . .	8
5.2. Artificial Network . . . . .	9
6. Conclusions and Future Work . . . . .	11
7. References . . . . .	12

## 1. Introduction

DeviantART is the most important online community of user-generated artworks. Launched in 2000 to entertain, inspire and empower the artist in all of us, today deviantART (commonly abbreviated as dA) is one of the largest online communities showcasing user-made artworks. With its 15 million members, 100 million images and 35 million unique visitors per month, surpassing any real or virtual museum, dA offers a genuine virtual space for disseminating art.

Unlike *Flickr* or *Youtube*, which focus on photographs and videos respectively, dA hosts a variety of genres, offering and even enforcing a delicate category structure to its users. Thus, all artwork is organized according to a comprehensive category structure, from photographs to various digital and traditional art forms. Each member of the site has his/her own webpage featuring a gallery, a journal, a favorites and a watchers section, as well as a basic information box highlighting statistics such as number of visitors, comments and downloads etc. These statics build up the main evaluation system of the dA community; a member with a large number of visitors/comments or an artwork with large number of favorites is seen as successful. The information box also contains demographic data (gender/geographic location/age), and gives details about member status. This rich background information allows us to study the dynamics of dA through network analysis.

In this project, we highlight impacts of favorites on their watchers. The probability of an artwork is normally liked by a member is one over the total number of artworks. However, there are some interfere factors that can increase this probability such as favorites. A member follows another member because he/she probably likes his/her artworks and honors his/her ideas. Therefore, if followed member adds an artwork to his/her favorite list, his/her follower will be probably affected and he/she will also add this artwork. Since dA has very large member and artwork database, research is done on certain categories in which members follow each other especially for contents.

Second aspect of the research is determining the person who is the most influential in the category. An effect coefficient will be calculated by traversing each artwork in preselected category. Effect coefficient of an user specifies his/her ability to attract others to make favorite artworks that s/he liked. The algorithm to calculate the effect coefficients will be verified by an artificial network. Finally, determined effective deviants will be followed in the community to double check the conclusions and the algorithm.

## 2. Difference of Current Project

DeviantART allows emerging and established artists to exhibit, promote, and share their works within a peer community dedicated to art. dA is a highly interactive and dynamic community where artists have their own profile containing their artwork. Each artist can add other artists to the friends list to automatically receive updates (e.g. newly added artwork) about these artists. There has been research into this feature. However, although favorite and watcher relation is as important as friend relation in shaping the network, there is no research on it.

### 2.1. Why is Favorite - Watcher Relation important?

Artists can be informed by friends list because when there is an update in the gallery of the friend, automatic information message is delivered. Therefore, artists know that their friend added some artworks probably attractive to them and they should check the gallery. There is an external trigger to be liked for an artwork. In favorite - watcher relation, followed artist can surf in several galleries to find interesting and attractive artworks but when s/he added the artwork to his/her favorite list, there would be no information message for his/her watchers.

### 2.2. How can Favorite - Watcher Relation shape the network?

There is a general tendency in people such that if people are attracted by someone and choose him/her as idol, they want to know what s/he is doing and check information sources regularly. This project is dealing with this generalization. Exceptions will also be determined such that if this hypothesis is wrong, the claim that nobody can affect others may also be wrong. There can be leaders in the group who can rapidly increase favorite number of artworks.

### 3. Data Fetch

Data gathering is a difficult and time-consuming since there is no provided application programming interface (API) by dA. This limits the number of studies and analysis on dA by complicating data gathering as our case or by forcing to be drawn the study.

We needed historical data to process the effects and there are three ways to gather data.

#### 3.1. Wayback Machine

*archive.org* is a non-profit digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, *archive.org* provides free access to researchers, historians, scholars and the general public.

However, this path is closed because although there are 1329 different pages, there isn't any complete snapshot of the dA. For example, just home pages are saved, user information isn't saved since there are accessible today but today information can't give the picture of the past because user information is very changeable and it changes in daily basis. Saved yearly data jumps between months, there are just few pages in each month.

#### 3.2. Periodic Snapshots

Second alternative is taking snapshots of dA regularly. We are just needed to fetch main browse directory of dA (*browse.deviantart.com*). However, to analyze the effects, we must create meaningful timeline which will approximately occupy 1 - 1.5 month. Coding phase is easy but delay to get data is unacceptable.



### 3.3. Intelligent Crawler

Third and last alternative is going along the difficult path, writing an intelligent crawler. Coding phase of the crawler is difficult since it must surf between web pages which have different formats (html codes) and regular expressions require a lot of care. Firstly, main browse directory (*browse.deviantart.com*) is visited as in snapshots case. Then, individual pages of deviants and deviations are visited to formulate the past data and matching properties.

Since the number of the visited and processed web pages is excessively increased, fetching time is also increased very much, relatively. Increased fetching time and huge database of artwork require limiting categories. We tried to choose a category which is manageable and in which deviants have a strong relationship to see the effect of the favorite relation easily.

Although coding phase is longer than snapshots crawler because of increased complexity, waiting time for crawler to complete is dramatically smaller. We have experienced about three days that is approximately 10% of the snapshots case.

## 4. Effect Coefficient - Algorithms

Second phase is composed of creating favorite change statistics per artwork. Six algorithms are tried which are based on three parameters. Namely, *fixedTotal*, *fixedTimeTotal*, *fixedTotalAverage*, *fixedTimeTotalAverage*, *weightedTotal*, *weightedTimeTotal*.

In *fixedTotal*, deviants are compared according to totally how many deviants have added the artwork that s/he had added before. Followed deviant adds one deviation to his/her list and then other deviants also add the same deviation to their lists. We count the number of deviants who are adding and comparison is applied on this number. The larger is the number, the more effective deviant means.

In *fixedTimeTotal*, deviants are compared according to time scaled *fixedTotal*. We calculate the time span by subtracting the time of the followed deviant from the time of the last deviant who has added the deviation. To adjust increase data according to time, we divide the number which is calculated as in the *fixedTotal* by time span.

In *fixedTotalAverage*, deviants are compared according to the average of *fixedTotal*. Denominator of algorithm is the number of deviations that this deviant has added to his/her favorite list.

In *fixedTimeTotalAverage*, deviants are compared according to the average of *fixedTimeTotal*. Denominator of algorithm is the number of deviations that this deviant has added to his/her favorite list.

In *weightedTotal*, deviants are compared according to weighted version of *fixedTotal*. Weights are calculated by dividing the number of deviants that has added this deviation to their list via total number of deviants.

In *fixedTotal*, deviants are compared according to weighted version of *fixedTimeTotal*. Weights are calculated by dividing the number of deviants that has added this deviation to their list via total number of deviants

## 5. Results of Applied Algorithms

3d game category is chosen for investigation because of its size, compact and content-based structure. Moreover, games are addictive and forward deviants to watch community regularly and this provides close relationships.

Since most deviants in this category are fans of some games, we hope that some deviants should have an important effect on others because if the deviant is proficient in this game, s/he will be an authority, for example he has done the highest score on the game. However, none of the algorithm gave satisfying results. In all algorithms, effectiveness of the deviants are distributed uniformly and we couldn't find really deviating user. This distribution probably results from smooth lines because the network cannot be totally random, there should be hierarchy between deviants as in the real life. We calculated effective coefficients from the whole life cycle of the deviations. If d1 is the authority and makes the difference, d2 is also affected and d2 also takes some effectiveness points from d1. In the same manner, so d3 and d4, etc. As a result of accumulation, derivatives come closer and when we calculate derivatives for all deviations, random output is generated. For one deviation d1 is the most effective since d1 added first and in the other one d2 is the most effective since d2 has added before than others.

### 5.1. Improvements

Since caring the whole life cycle brings averaging on derivatives, we focused on dividing the life cycle into subfields and comparing the deviants according to the performance of deviants in the this time span. In this version, averaging behavior continue to be dominant since the data is continuous but the effect diminished a bit.

## 5.2. Artificial Network

Formulating a lot of algorithm to find the most effective deviant and applying each algorithm on actual data is time consuming since actual data is relatively large. Moreover, we are't sure about correctness of the algorithm. Therefore, we created an artificial network. Applying the algorithm on this network is easy and efficient since we control the network and can change the parameters to what we want. Since we know the parameters, we can easily check the power of the algorithm.

There are two main objects in artificial network as in the real network, namely deviant and deviation.

Deviant has an *ID*, an *effectCoefficient* and a *favoriteCount*. *ID* is necessary to identify the deviant whether it is an authority or not. *effectCoefficient* comes from uniform distribution except the authorities we assigned and represents the power of the deviant to attract other deviants to choose what s/he chooses. *favoriteCount* comes from normal distribution and represents the the number of deviations will be selected by this deviant.

Deviation has an *ID*, a *qualityCoefficient*, a *favoriteList* and a *favoriteTimes*. *ID* is used as in deviant. *qualityCoefficient* represents how this artwork is good or attractive independent of deviants that has added it to their list and comes from uniform distribution. *favoriteList* is the deviant list that have added this deviant to their favorite list and *favoriteTimes* is time list that holds the adding times.

To generate network, firstly we generate deviants and deviations. Moreover, there are four effective deviants which have 0.9, 0.7, 0.5, 0.3 effect coefficients, respectively.

Generation algorithm starts with choosing a random deviant. Then, algorithms chooses the most popular deviation with 0.25 probability, or a deviation from the list of authorities (first effective with 0.4 probability, second 0.3, third 0.2 and fourth 0.1) with 0.25 probability, or a totally random deviation. Then, if the list of the deviation is empty, deviant chooses it as favorite with  $0.75 * \text{its quality coefficient}$ , otherwise select is made according to max of its quality coefficient and max of effect coefficients of deviants in the list. Moreover, this procedure repeats until *favoriteCount* of all deviants become zero.

## 6. Conclusions and Future Work

Until now, familiarity with dA is accomplished by creating an account and using features and analyzing the html code. An intelligent crawler is written to gather the data. Some algorithms are tested on actual data but they didn't give good results so we trying to improve it. Moreover, an artifical network is created to easily check the power of the formulated algorithms.

Some data is coming from dA. We will analyze it and improve our algorithm and apply it on actual data and determine some deviants as authorities. Moreover, we will follow determined deviants for some period in dA to double check the findings.

## 7. References

- <http://www.deviantart.com>
- **DeviantART in Spotlight: A Network of Arts and Artists**  
Almila Akdag Salah, Albert Ali Salah, Bart Buter, Nick Dijkshoorn, Davide Modolo, Quang Nguyen, Sander van Noort, Bart van de Poel
- **TADA: Toolkit for Analysis of deviantART**  
Bart Buter, Nick Dijkshoorn, Davide Modolo, Quang Nguyen, Sander van Noort, Bart van de Poel
- **Community Structure in social and biological networks**  
M. Girvan and E.J Newman
- <http://en.wikipedia.org/wiki/DeviantArt>
- <http://www.dmusic.com>
- [http://web.archive.org/web/\\*/http://www.deviantart.com](http://web.archive.org/web/*/http://www.deviantart.com)