



Fan History LLC  
900 Glen Oak Drive  
Sleepy Hollow IL 60118  
<http://www.fanhistory.com>

## Fan Fiction's Predictive Value For Nielsen Ratings

*By Laura Hale,  
Fan History LLC Founder  
<http://www.fanhistory.com/wiki/User:Laura>  
[laura@fanhistory.com](mailto:laura@fanhistory.com)*

## Contents

---

<b>ABSTRACT.....</b>	<b>3</b>
<b>INTRODUCTION.....</b>	<b>4</b>
<b>DECONSTRUCTING THE DATA BACK STORY.....</b>	<b>5</b>
DATA COLLECTION METHODOLOGY.....	5
TELEVISION SHOWS IN SAMPLE.....	6
FAN FICTION DAILY POSTING VOLUME TIME PERIOD.....	11
<b>CONSTRUCTING MEANING FROM THE DATA SET.....</b>	<b>12</b>
GENERAL PATTERNS.....	12
NETWORK PATTERNS.....	14
GENRE PATTERNS.....	15
PATTERNS BY SHOW.....	17
<b>CONCLUSION.....</b>	<b>21</b>

## Fan Fiction's Predictive Value For Nielsen Ratings

---

*Laura Hale*

*Fan History LLC, Dundee, Illinois*

On January 15, 2009, CSI had one of its highest rated episodes all season. On that day, people published 26 new pieces of fan fiction, the most stories posted on the same day as an episode had aired. On September 25, 2008, CSI had its third lowest ratings day all season and people posted zero new stories on that date.

Fan fiction is a really popular outlet for fan expression of interest in television shows. The stories are creative, explore plot lines in the show and, according to many fans, help market a series in a positive way. Fans often argue that their activities mirror larger interest in a show, and that producers should pay more attention to them and cater to their fanish interests as the example provided seems to demonstrate. Fan Fiction's Predictive Value for Nielsen Ratings tests this fan theory and answers the question: Does the volume of fan fiction published in the period around when an episode airs correlate to Nielsen Ratings?

To answer this question, fan fiction daily posting stats were gathered for the one week period around television shows where fan fiction communities existed and Nielsen Ratings were available for that show. The fan fiction data was compiled from six archives: FanFiction.Net, fanfiktion.de, FanWorks.Org, FicWad, SkyHawke, and Freedom of Speech Fan Fiction. The Nielsen Ratings data included over 720 episodes representing thirty-nine shows. Once this data was compiled, it was analyzed using Pearson's Correlation and linear regression.

The results confirmed what many fans already suspected: Levels of fan activity, specifically in terms of the production of fan fiction, mirrors interest specific episodes of television. Fan fiction can be used to predict Nielsen Ratings. The predictive value is strengthened in several cases when it is broken down by network, genre or specific television show. The best networks for predicting Nielsen Ratings are CBS, The CW, Disney, Fox and USA. Comedy, crime comedy, crime drama, medical comedy and sports drama are the best genres for predicting Nielsen Ratings. The strongest correlations for television shows for predicting Nielsen Ratings are Burn Notice, CSI, Eli Stone, Friday Night Lights, Gossip Girl, Grey's Anatomy, Hannah Montana, Heroes, iCarly, Law and Order: Criminal Intent, Life, Prison Break, Psych, and Terminator: Sarah Connor Chronicles.

This information is potentially valuable to parties with a vested interest in a television show's performance. By analyzing content patterns around periods with high volumes of fan fiction and high Nielsen Ratings, comparing that to periods of low posting volume and lower Nielsen Ratings, producers can make changes to maintain high interest amongst fans. Non-American television networks and advertisers can better predict how their shows will perform. This method of analysis can help organizations save money as it is cheaper to monitor and track than other analytic tools.

## Introduction

On January 15, 2009, CSI had one of its highest rated episodes all season. On that day, people published 26 new pieces of fan fiction, the most stories posted on the same day as an episode had aired. On September 25, 2008, CSI had its third lowest ratings day all season and people posted zero new stories on that date.

Fan fiction. Lots of people write it in response to their favorite television shows, movies, video games, books and bands. The largest site for it, FanFiction.Net, gets over a million visitors from the United States every month. There are large communities of fan fiction writers and readers on other popular sites like LiveJournal and Quizilla. There are thousands of smaller sites, message boards, blogging communities and mailing lists that get several thousand visitors a month. The fan fiction community is large, active and dedicated to the object of its fascination.

The general assumption by fans is that their works, like fanfiction, mirror popular interest in a product by a larger community. Some fans feel that their stories are a form of word of mouth marketing that helps generate additional interest in television shows. Given that, they often feel entitled to attention from producers, where the producers should cater to their specific interests in certain story lines and relationships. This is a theory that generally has not been explored. Little market research has been accomplished to determine the validity of these underlying assumptions by fans as they would be applicable to content producers and industries reliant upon them.

Lots of money is spent promoting episodes to help gain a larger audience share. Lots of money is also spent by advertisers who are seeking to promote their products to the audience watching those shows. There are currently a number of tools out there that can help predict potential market reach and audience demographics before a television show goes on the air. Most of the research after that is largely based on previous performance for a show using Nielsen Ratings. Given the long history of use for these existing tools, there is a substantial risk in switching analytic tools to try to ascertain market reach for a particular television show. It can be costly, and if the method is largely untested, lead to poor decision making.

This paper uses one analytic method to determine the validity of this fan assumption as a decision-making tool. It puts to the test the assumption in the fan community that their works mirror interest expressed in a show by tuning in and watching it at its scheduled time. It does this by trying to discover if there is a predictive value for fan fiction posting levels and Nielsen Ratings by looking at correlation between the two variables.

## **Deconstructing the data back story**

### *Data collection methodology*

Data from this study includes two types: fan fiction posting volume and Nielsen Ratings. Data for fan fiction posting volume came from Fan History's fan fiction community size articles. ([http://www.fanhistory.com/wiki/Category:Fan\\_fiction\\_community\\_size](http://www.fanhistory.com/wiki/Category:Fan_fiction_community_size)) Data for Nielsen Ratings was collected from two sources: Nielsen Wire's Media and Entertainment blog ([http://blog.nielsen.com/nielsenwire/media\\_entertainment/](http://blog.nielsen.com/nielsenwire/media_entertainment/)) and TVbytheNumbers.com (<http://tvbythenumbers.com/>). The data set represents a period between September 12, 2008 and July 8, 2009.

Fan fiction daily posting volume data was collected from Fan History's fan fiction community size articles. The data on Fan History was obtained on a daily basis using a bot that took raw data from a selection of six fan fiction archives, formatting it and outputting it in wiki form. The selection of television shows represented on Fan History was chosen in September 2008 by compiling a list of all the fandoms represented on the following fan fiction archives: FanFiction.Net, fanfiktion.de, FanWorks.Org, FicWad, SkyHawke, and Freedom of Speech Fan Fiction. The list of shows was not updated during this period to reflect new shows that started to air. Each day between September 10, 2008 and July 8, 2009, the bot would visit each site, record the total number of stories for a category, record that number, compare it to the total number of stories published the previous day and record the difference. The total for all new stories across the five archives in the sample were added together to arrive at one number for how many pieces of fan fiction were posted on a given day.

In collecting Nielsen Ratings, the primary collection source was Nielsen Wire's Media and Entertainment blog. When Nielsen Ratings for a show could not be found using that resource, TVbytheNumbers.com was used instead. Nielsen ratings were only included for reruns and original runs of episodes in their scheduled primetime slot on their home network. We only collected Nielsen Ratings for shows where we had data regarding daily fan fiction volume.

*Television shows in sample*

The following television shows were represented in the sample:

24  
Bones  
Boston Legal  
Burn Notice  
Chuck  
Criminal Minds'  
CSI  
CSI: Miami  
CSI: New York  
Desperate Housewives  
Eli Stone  
ER  
Friday Night Lights  
Gossip Girl  
Grey's Anatomy  
Hannah Montana  
Heroes  
House M.D.  
iCarly  
Knight Rider  
Law and Order  
Law and Order: CI  
Law and Order: SVU  
Life  
Life on Mars  
Lost  
Medium  
My Name is Earl  
NCIS  
Numb3rs  
Prison Break  
Private Practice  
Psych  
Pushing Daises  
Scrubs  
Smallville  
Terminator: Sarah Connor  
Chronicles  
The Office  
Ugly Betty  
Without a Trace

These shows fall into several general broad categories. Shows can be sorted based on the network they air on: ABC, NBC, CBS, Disney, FOX, CW, and USA. They can also be

classified by genre: Dramas, science fiction and fantasy, crime shows, sitcoms, medical and family.

Some genres and networks seem to lend themselves to having a smaller fan fiction communities based around them. While there are more shows that ranked in the top ten on the Nielsen ratings, many did not have fan fiction communities based around them.

One such genre is reality television. There is fan fiction based on some shows like So You Think You Can Dance but the community is significantly smaller and organizes around show specific message boards. Fan fiction in that community is often posted alongside other content, like episode reviews and role plays, where it is difficult to differentiate between content.

Other genres such as sitcoms lack fan fiction communities. Sometimes this is a result of a show lacking a strong relationship for fans to obsess over. Some of it is a result of characters not having much depth for writers to explore. The lack of a larger community for sitcoms also happens because episodes are self-contained; you do not need knowledge of the whole body of canon to sit down and enjoy a show. This means there is not necessarily the need for writers to explore characters in depth. The problem for sitcoms is also compounded because writers know the audience for their work is small so there is less incentive to publish fan fiction for those shows.

Family shows are also under represented in this sample. Shows in this genre lend themselves a bit more to fan fiction than sitcoms so the fan fiction community is larger. The audience tends to be younger. As a result, the fan fiction community tends to congregate around sites that allow less adult content and are tailored for their age group.

Sporting events occasionally appeared on the top of the Nielsen ratings for a time period. There are communities for sports based fan fiction but they are not found on FanFiction.Net or most multifandom fan fiction archives. The material is viewed as offensive by some members of the larger sports fandom community because it often depicts athletes as being involved in same sex romantic or sexual relationships. Fans worry about getting in to legal trouble for publishing that sort of fan fiction as there is precedent for this. One case involved a legal threat from Andy Pettite to FanDomination.Net for writing him as homosexual in several stories. Another case involved a female fan writing same stories involving members of the Ohio Buckeyes women's basketball team; the university responded by legally threatening her and removing her booster status. There have been instances where athletes or teams have threatened lawsuits for this material. It has pushed it underground. Given that, no sporting events are included in this sample.

There are sizable fan fiction communities based around soap operas and cartoons. Soaps were not included in this sample because Nielsen ratings for shows like General Hospital and Days of Our Lives were not available. Cartoons had similar issues: Nielsen ratings were not available and most shows were not in original runs on their homenetworks.

Genres that tend to attract large fan fiction communities tend to be dramas, crime dramas, medical dramas and science fiction shows. All of these genres have long histories of fan fiction communities surrounding specific shows, some of which date back to the late 1960s and early 1970s.



### *Fan fiction archives in sample*

There were six archives represented in the sample. These archives were chosen for inclusion on Fan History's original data set because they represented a large cross section of the fan community around the world, represented a large number of fandoms, catered to different niches and did not always overlap in their core audience. They included FanFiction.Net, fanfiktion.de, FanWorks.Org, FicWad, Freedom of Speech Fan Fiction and Skyhawke.

FanFiction.Net is the most popular fan fiction archive on the Internet. The only fan fiction archive that comes close to it in volume and traffic is AdultFanFiction.Net and they get an eighth of the amount of FanFiction.Net. According to Quantcast, for July 2009, the site was 892<sup>th</sup> most visited site by Americans. The site caters to a mostly American audience, though it has a large international contingent who publish stories in languages other than English. In this sample, non-English stories were counted along with English stories as there was no way to eliminate them. Of the archives in the sample, this site represented roughly ninety-nine percent of the daily activity for most fandoms. It also represented every fandom included on the list.

fanfiktion.de is one of the most popular fan fiction archives in Germany and Austria. According to Quantcast, for July 2009, the site was ranked 767,620. It has more traffic than the Quantcast number suggests because the site is in German and caters to that audience. Alexa ranks it as the 1,204<sup>th</sup> most popular site in Germany and the 2,527<sup>th</sup> most popular site in Austria. All the stories are in German. Of the archives in the sample, fanfiktion.de represented the second largest amount of stories. It was included because other language communities frequently are privy to American spoilers as a result of subbing and bilingual speakers in the community. Their interest in episodes in terms of fan fiction is often reflective of the English-speaking community and they will often write stories based on spoilers, descriptions of episodes or of torrents posted soon after an episode airs in the United States.

FanWorks.Org and FicWad are both small archives with a small, dedicated audience. According to Quantcast, for July 2009, FanWorks.Org was ranked 178,607 and FicWad was ranked 80,760. They were more popular a few years ago than they are now. This is in some small part due to absentee landlords who do not actively publicize the site or update the code. Both are more popular and well-known for other fan fiction communities like anime, actors and music. The television communities are small and not very active in comparison. Both sites tended to draw heavily from communities not attached to FanFiction.Net. FicWad tended to draw from LiveJournal and FanWorks.Org drew from Quizilla.

Freedom of Speech Fan Fiction is an archive in trouble. According to Quantcast, for July 2009, the site was ranked 84,054. This is despite the fact that it faced intense downtime for a while during late July and most of August. It had a hard core user base mostly composed of anime writers. Its television section is small and there are only one or two active fandoms. The community on Freedom of Speech Fan Fiction currently wants the site to become stable once again but its future is questionable. If the site were to close, the volume of stories posted if this site were removed from the sample would be negligible and not affect the results in a meaningful way.

Skyhawke is an Australian-based fan fiction archive and subject to Australian laws. Given this, it caters to a mostly Australian audience who are worried about running afoul of some of their domestic laws that United States-based archives may not be compliant with. The site also caters to Australians who are interested in not being spoiled for episodes before they are broadcast in Australia. Quantcast does not rank this site but estimates it gets roughly 12,400 US based visitors a month. Alexa does rank this site and puts it at 83,379<sup>th</sup> and it is ranked 64,508<sup>th</sup> in Australia.

All of these archives have some fundamental organizational and policy principles. They generally organize stories around specific fandoms; they don't organize around characters, ratings, genres or authors. This organizational structure is based around a category, where the top-level category shows you the number of stories published in subcategories. That means that it is possible to track the volume of fan fiction posted. This was one of the key reasons for choosing these archives. Sites like Quizilla and LiveJournal have large fan fiction communities but there is no easy means of calculating how many stories are posted on a daily basis. Other sites like FanDomination.Net did not update their story totals in their categories in real time; rather totals are updated manually. The policy principle that these archives also share is that they allow users to remove their content when they chose to. What this means in terms of the data set that is generated is that there are frequently days where the posting total is negative; people can and do take advantage of this.

### *Fan fiction daily posting volume time period*

The window of data for including daily fan fiction posting volume and its possible predictive value was three days before to three days after an episode aired. This time period was chosen for several reasons based around episode promotion and established patterns in the fan fiction community.

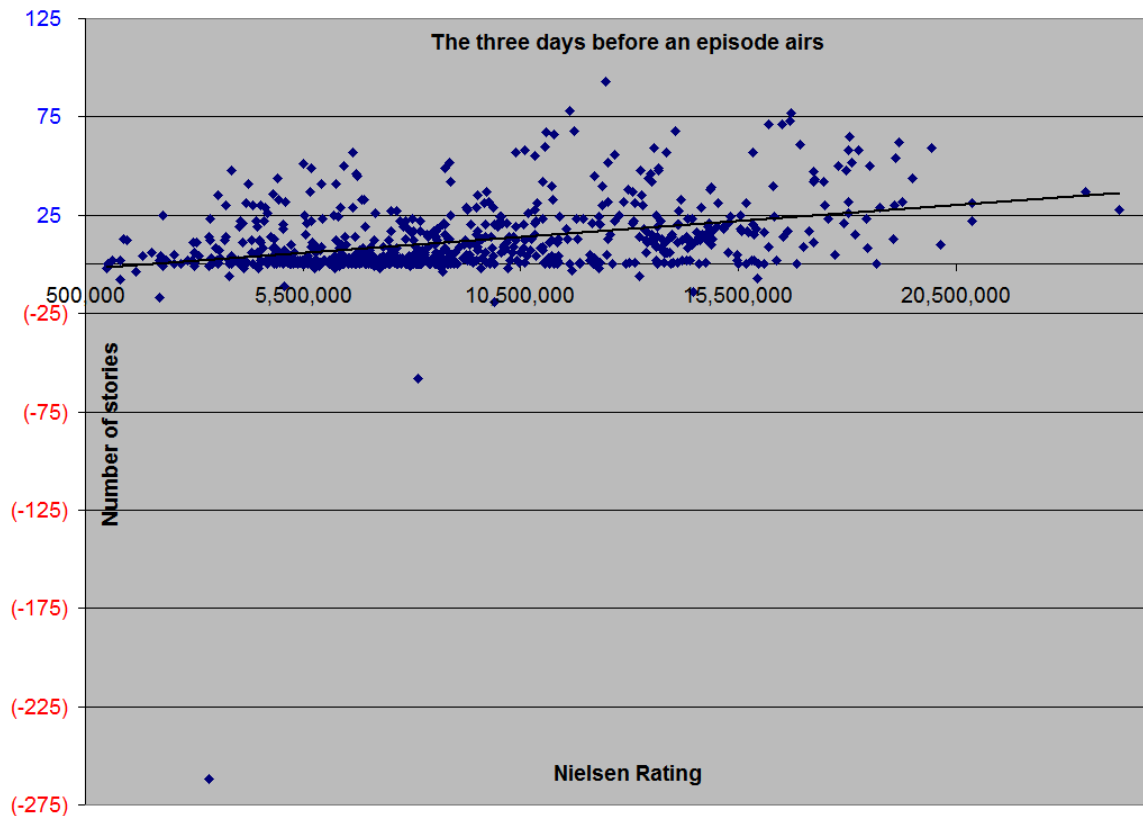
Most episodes of television are not promoted more than a week out. The major exceptions are for season premieres or after a show returns from hiatus. There is no reason to assume that going back any earlier would result in any predictive value for an episode.

In the fan community, there is a long tradition of writing post episode stories and publishing them within a few days of the episode originally airing. When deciding the sampling period to see if you can predict the ratings for an episode, this was taken in to account. The assumption was made that post episode stories were generally written three days after an episode was aired. To account for this, only stories published three days before an episode were looked at in order to try to predict Nielsen ratings based on fan fiction posting volume. The assumption is that there is not bleed over factor for post episode fan fiction beyond the three days after an episode aired.

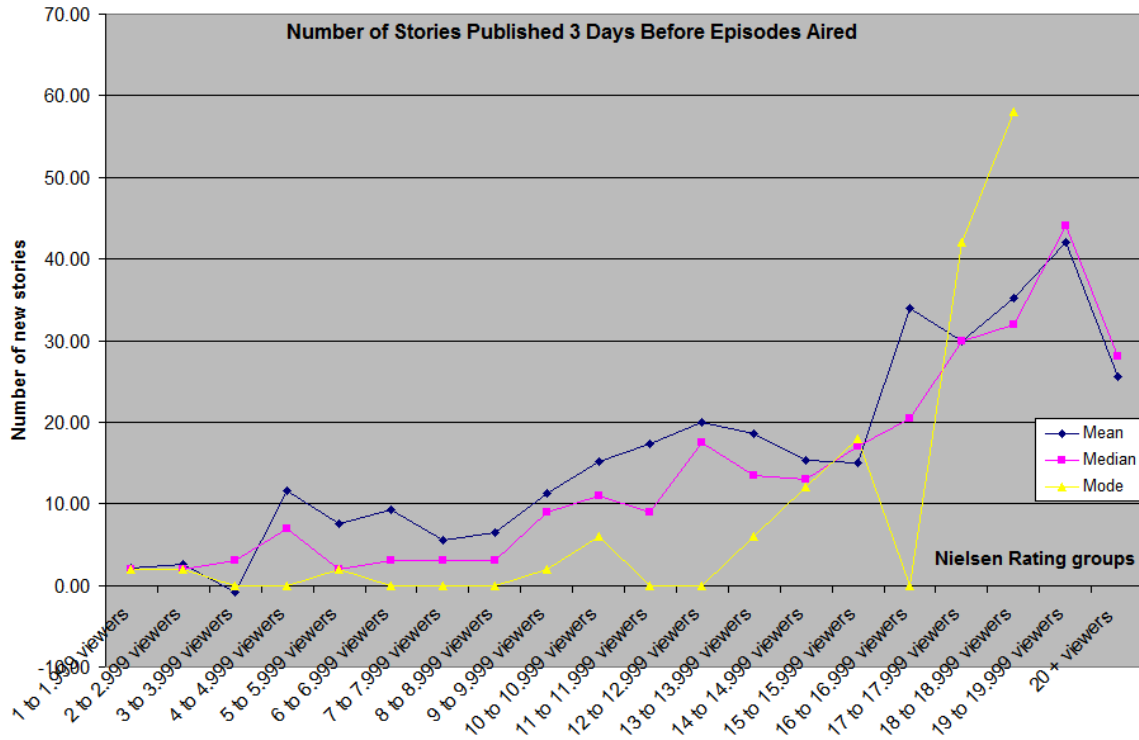
## Constructing meaning from the data set

### General patterns

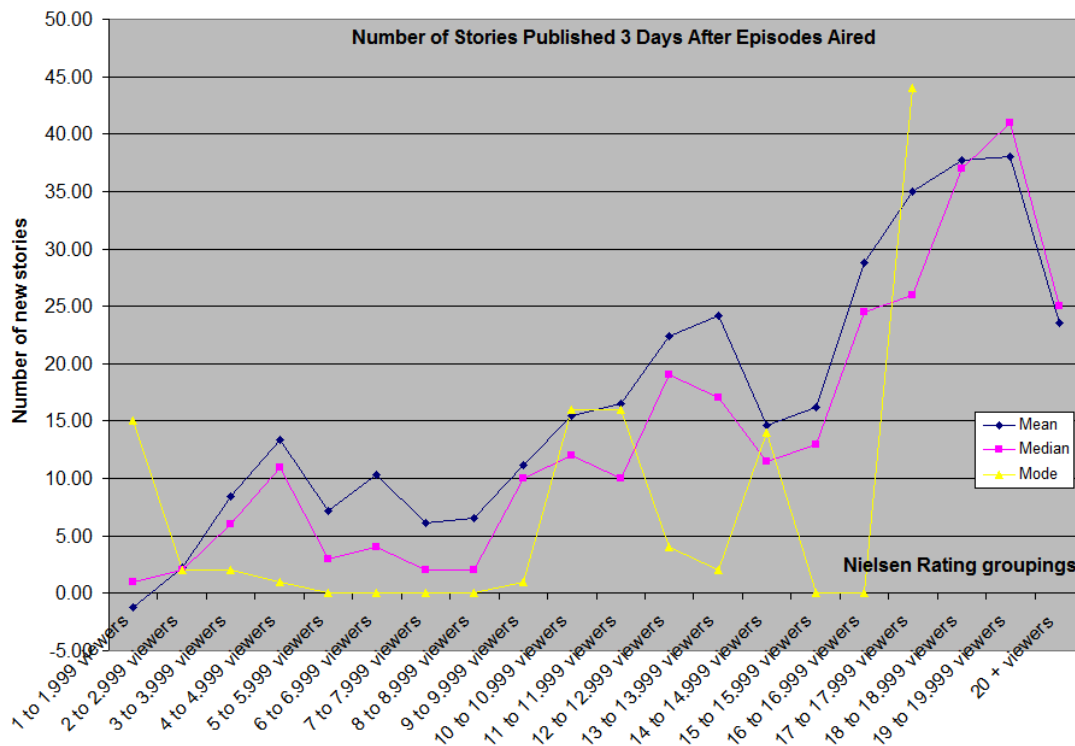
Fan fiction can be used as a nominal predictor of Nielsen Ratings. The relationship is not a straight line, where it can be said that if there are X amount of stories published, the Nielsen Ratings will be Y. Still, there is a strong enough correlation, with enough points grouped together, that you can make some good assumptions about future Nielsen Ratings based on fan fiction posted prior to an episode airing.



When the Nielsen Rating is lower, generally the posting volume in the three day period before an episode airs is lower. When the Nielsen Rating is higher, the posting volume is generally higher. No episode that had a Nielsen rating of 18,745,000 or higher had fewer than 10 new stories published in that period. This pattern is clear when looking at mean, median and mode for new stories based around Nielsen Rating groupings.



The pattern of seeing an increase in Nielsen Ratings for increased fan fiction posting volume is also present for the three day period after an episode airs. The correlation is .3867, higher than the preceding three day period which has a correlation of .3533. When broken down in to Nidsen Ratings groups, the pattern is clearly present.



### Network patterns

The network that episodes air on can influence the predictive value of fan fiction published prior to an episode air. CBS, (Criminal Minds, CSI, CSI: Miami, CSI: New York, NCIS, Numb3rs, Without a Trace) NBC (Chuck, ER, Friday Night Lights, Heroes, Knight Rider, Law and Order, Law and Order: Criminal Intent, Law and Order: Special Victims Unit, Life, Medium, My Name is Earl, The Office) and USA (Burn Notice, Psych) are the three networks with the strongest correlation between Nielsen Ratings and fan fiction posting volume.

Network	3 Before	2 Before	Before	Of	After	2 After	3 After	C Before	C After
All	<b>0.3024</b>	<b>0.3184</b>	0.2913	<b>0.3577</b>	<b>0.3136</b>	<b>0.3331</b>	<b>0.3648</b>	<b>0.3534</b>	<b>0.3868</b>
ABC	0.1978	0.2393	0.0716	0.1867	0.1832	0.2344	0.2682	0.2144	0.2687
CBS	<b>0.3245</b>	<b>0.3933</b>	0.2997	<b>0.3929</b>	<b>0.3567</b>	<b>0.3625</b>	<b>0.3760</b>	<b>0.3942</b>	<b>0.4274</b>
CW	<b>0.4296</b>	-0.0650	-0.0521	0.1788	0.0758	<b>0.3420</b>	<b>0.3469</b>	-0.0174	<b>0.4037</b>
Disney	<b>0.3107</b>	0.1446	0.0817	-0.0667	-0.2255	0.1611	0.2497	0.2746	-0.0107
Fox	0.0832	0.1844	0.2379	0.2635	0.2433	0.1756	<b>0.3208</b>	0.1946	0.2934
NBC	0.1790	0.2491	0.2515	0.2217	0.2599	0.2521	0.2065	<b>0.3089</b>	0.2800
USA	-0.2981	-0.1951	-0.1975	<b>-0.4681</b>	<b>-0.3562</b>	-0.2361	<b>-0.3360</b>	<b>-0.3269</b>	<b>-0.3443</b>

In the case of CBS and NBC, the relationship is a positive one: When Nielsen ratings are higher in the three-day period before an episode airs, more fan fiction is posted. USA is the opposite: The higher the Nielsen Rating, the lower the number of stories that are published.

On individual days prior to an episode airing, CBS, CW (Gossip Girl, Smallville) and Disney (Hannah Montana, iCarly) had a meaningful correlation. This correlation happened three days before an episode aired. All the meaningful correlations on the individual days were positive: Increase in Nielsen ratings saw an increase in fan fiction publishing levels.

In the three days after an episode air, there are more meaningful correlations than in the three days preceding. Networks with meaningful correlations are CBS, CW and USA. Fox (24, Bones, House M.D., Prison Break, Terminator: Sarah Connor Chronicles) is close to being meaningful, as is ABC, (Boston Legal, Desperate Housewives, Eli Stone, Grey's Anatomy, Life on Mars, Lost, Private Practice, Scrubs, Ugly Betty) especially when compared to the preceding period.

### Genre patterns

When shows are sorted by genre, more patterns begin to emerge that can, in some cases, help to better predict Nielsen ratings. These genres are comedy (My Name is Earl, The Office), crime comedy (Psych), crime drama, (Bones, Criminal Minds, CSI, CSI: Miami, CSI: New York, Law and Order: Criminal Minds, Law and Order: Special Victims Unit, Life, Medium, NCIS, Numb3rs, Prison Break, Without A Trace) and sports drama. (Friday Night Lights) Some genres have no correlation, where the relationship between fan fiction posting volume in the three days before and Nielsen Rating is totally random. These genres include dramatic comedy, (Chuck, Desperate Housewives, Gossip Girl, Ugly Betty) legal comedy, (Boston Legal) legal drama, (Eli Stone, Law and Order, Private Practice) medical comedy (Scrubs), medical drama, and thriller. (24, Burn Notice) The two remaining genres, science fiction and fantasy, (Heroes, Knight Rider, Life on Mars, Lost, Pushing Daisies, Smallville, Terminator: Sarah Connor Chronicles) and family (Hannah Montana, iCarly) have some meaningful relationship but are a bit closer to being random.

Some genres with little correlation for the three days before have days during that period where there is a stronger, more meaningful correlation. For family and medical comedy, three days before an episode airs is important. Family shows an increase in the posting volume for episodes with higher Nielsen Ratings. Medical comedy shows the opposite, where a higher rated episode sees a decrease in the number of stories posted on the third day out. Comedy, crime comedy, crime drama follow the same single day pattern three days out as family. Two days before an episode airs, thriller sees an increase in posting volume that mirrors an increase in Nielsen Ratings. Crime comedy and crime drama share that pattern. One day before an episode airs, medical comedy and sports drama have a positive correlation between fan fiction posting volume and Nielsen Ratings.

**Pearson's Correlation between fan fiction posting and Nielsen Ratings**

	3 Before	2 Before	Before	Of	After	2 After	3 After	C Before	C After
All	<b>0.3024</b>	<b>0.3184</b>	0.2913	<b>0.3577</b>	<b>0.3136</b>	<b>0.3331</b>	<b>0.3648</b>	<b>0.3534</b>	<b>0.3868</b>
Comedy	<b>0.4302</b>	0.2968	<b>0.3287</b>	<b>0.3094</b>	0.0997	<b>0.3017</b>	0.1212	<b>0.5278</b>	0.2352
Crime comedy	<b>0.9970</b>	<b>0.6515</b>	-0.2334	<b>0.9583</b>	<b>0.3525</b>	<b>0.9535</b>	0.0724	<b>0.9166</b>	<b>0.7293</b>
Crime drama	<b>0.3389</b>	<b>0.4160</b>	<b>0.3481</b>	<b>0.4178</b>	<b>0.3863</b>	<b>0.3616</b>	<b>0.3953</b>	<b>0.4280</b>	<b>0.4399</b>
Dramatic comedy	-0.2105	0.1172	0.0940	-0.0854	0.0451	-0.2056	0.0872	0.0884	0.0111
Family	<b>0.3107</b>	0.1446	0.0817	-0.0667	-0.2255	0.1611	0.2497	0.2746	-0.0107
Legal comedy	#DIV/0!	#DIV/0!	#DIV/0!	<b>-0.3129</b>	-0.2261	0.1343	0.1343	#DIV/0!	-0.0684
Legal drama	0.0504	0.0334	0.0877	0.0501	0.1120	0.0615	0.0072	0.0815	0.0948
Medical comedy	<b>-0.3927</b>	-0.0780	<b>0.4047</b>	0.0135	<b>0.4209</b>	-0.1060	0.2439	0.0198	<b>0.4110</b>
Medical drama	0.0179	0.0418	-0.0577	0.0065	0.0467	0.0080	0.1331	-0.0048	0.0735
Science fiction and fantasy	0.0060	0.1485	0.2867	0.2049	-0.0282	0.2221	0.0625	0.1564	0.1291
Sports drama	-0.0997	-0.0672	<b>0.8069</b>	<b>0.3315</b>	-0.1342	0.0900	-0.0347	<b>0.4632</b>	-0.0282
Thriller	-0.2019	<b>0.3271</b>	0.1508	<b>0.3584</b>	0.0976	-0.0396	0.2351	0.0567	0.1459

Fan fiction volume posted day of is the period with the most genres showing predictive value. This is a bit misleading as three of these genres are represented by only one show: Crime drama with Psych, legal drama with Boston Legal and sports drama with Friday Night Lights. With the exception of legal comedy, all genres with significant meaning have a positive trend.

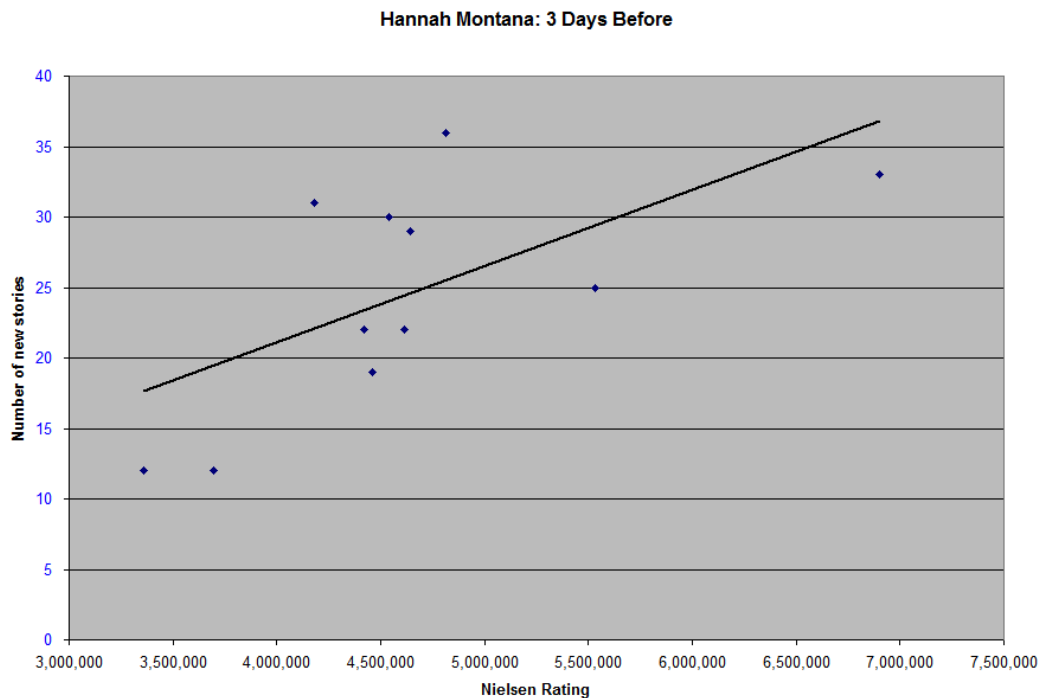
In the three day periods after an episode airs, the overall correlation is stronger than the three day periods before an episode airs. However when broken down by genre, fewer genres are predictive post episode airing than before an episode airs. Genres, with the exception of crime drama and maybe crime comedy and medical comedy, are not good predictors of the amount of fan fiction produced after episodes air based on the Nielsen Ratings.



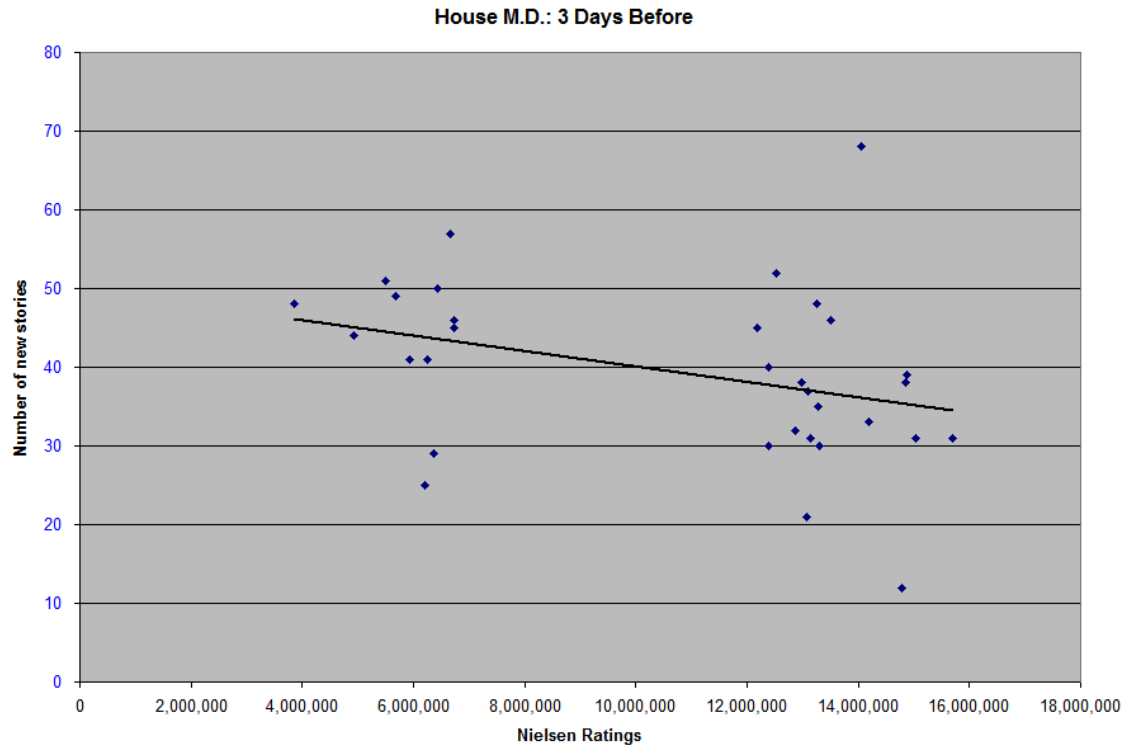
### Patterns by show

Different shows behave differently. Some have unique patterns all their own that are independent of the norms for the whole of the sample. Most do not deviate though. Where they might deviate, it is only to be a null correlation, where there is no relationship between fan fiction posting volume and Nielsen Ratings.

In the three day period before an episode airs, Hannah Montana, Law and Order: Criminal Intent, Prison Break, Psych and Terminator: The Sarah Connor Chronicles have the strongest correlation, where the more fan fiction that is posted, the higher the Nielsen Ratings. For Law and Order: Criminal Intent, Prison Break, Psych and Terminator: The Sarah Connor Chronicles, no show has a sample larger than eight episodes. In some cases, there are as few as four episodes representing the show. Law and Order: Criminal Intent suffered this problem. For this show, Nielsen did not mention some first runs of the show because reruns of episodes on USA had higher Nielsen Ratings than original episodes on NBC. Psych's potential audience was always smaller because it ran on cable and often competed with higher rated sporting events. Terminator: The Sarah Connor Chronicles had the issue of low ratings, not being rerun during weeks with new episodes and a cancellation in May 2009. The data says the correlation is meaningful but for those shows, the data set is too small to be entirely sure of the accuracy. Only Hannah Montana has a sample size where the results can be trusted.



Several shows have a meaningful correlation similar to that for the whole of the sample for this period. The correlation is between .3000 and .4999, where as the shows mentioned earlier have a correlation of .5000 or above. These shows include Friday Night Lights, Heroes, House M.D., Life, Medium, and The Office. All but House M.D. see an increase in the number of stories published when a show has higher Nielsen Ratings.

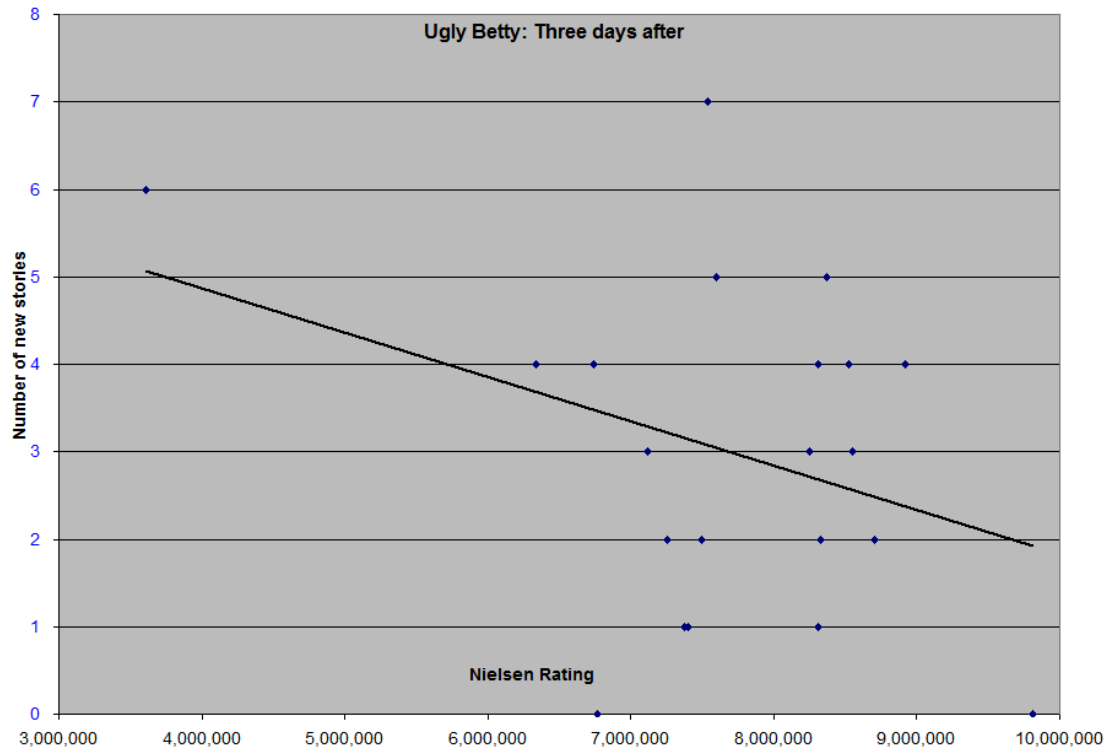


In the sample, there are many days when people removed stories and the number of new stories is negative. If it is assumed that the removal of stories is an independent variable that is irrelevant to actual posting, then those negative numbers can be set to zeros. When this is done, the results change and the correlation becomes stronger in several cases. Friday Night Lights moves from meaningful to very meaningful in a positive direction. Gossip Girl moved from random negative correlation to strong positive correlation. Hannah Montana, The Office and Prison Break's correlation became a bit more meaningful. Life on Mars went from random negative correlation to strong negative correlation. Private Practice went from random negative correlation to meaningful negative correlation.

Correlation after single day negative numbers are zeroed										Correlation after single day negative numbers are zeroed									
	3 Before	2 Before	Before	0	After	2 After	3 After	C Before	C After		3 Before	2 Before	Before	0	After	2 After	3 After	C Before	C After
All	0.3236	0.3702	0.3306	0.3601	0.3151	0.3428	0.3564	0.3867	0.3808	All	0.3236	0.3702	0.3306	0.3601	0.3151	0.3428	0.3564	0.3867	0.3808
24	-0.0257	-0.1585	0.2125	0.4889	0.0442	0.1891	0.4491	0.0801	0.3277	24	-0.0257	-0.1585	0.2125	0.4889	0.0442	0.1891	0.4491	0.0801	0.3277
Bones	-0.1495	-0.1631	-0.0596	0.2336	0.0629	-0.2482	-0.2139	-0.1680	-0.2015	Bones	-0.1495	-0.1631	-0.0596	0.2336	0.0629	-0.2482	-0.2139	-0.1680	-0.2015
Boston Legal	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.3129	-0.2261	0.1343	0.1343	#DIV/0!	Boston Legal	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.3129	-0.2261	0.1343	0.1343	#DIV/0!
Burn Notice	-0.1803	0.5940	-0.0446	-0.8106	-0.0828	-0.1218	0.1764	0.0517	-0.0311	Burn Notice	-0.1803	0.5940	-0.0446	-0.8106	-0.0828	-0.1218	0.1764	0.0517	-0.0311
Chuck	0.1114	-0.2431	-0.0758	-0.0910	-0.2440	0.0593	-0.0700	-0.1140	-0.1539	Chuck	0.1114	-0.2431	-0.0758	-0.0910	-0.2440	0.0593	-0.0700	-0.1140	-0.1539
Criminal Minds	-0.2584	0.2402	0.0328	-0.1619	0.0264	-0.1220	0.1004	0.0704	0.0128	Criminal Minds	-0.2584	0.2402	0.0328	-0.1619	0.0264	-0.1220	0.1004	0.0704	0.0128
CSI	0.2250	0.0785	0.0398	0.5560	0.4821	0.2820	0.1560	0.1896	0.4511	CSI	0.2250	0.0785	0.0398	0.5560	0.4821	0.2820	0.1560	0.1896	0.4511
CSI: Miami	-0.2066	0.1231	0.0064	-0.2992	0.1656	0.1144	0.3048	-0.0557	0.3229	CSI: Miami	-0.2066	0.1231	0.0064	-0.2992	0.1656	0.1144	0.3048	-0.0557	0.3229
CSI: New York	-0.2655	-0.2630	-0.0935	-0.2671	-0.1394	-0.0042	-0.0133	-0.2670	-0.1049	CSI: New York	-0.2655	-0.2630	-0.0935	-0.2671	-0.1394	-0.0042	-0.0133	-0.2670	-0.1049
Desperate Housewives	-0.2785	-0.2620	0.0188	-0.1403	0.1444	-0.3070	-0.2029	-0.2971	-0.2060	Desperate Housewives	-0.2785	-0.2620	0.0188	-0.1403	0.1444	-0.3070	-0.2029	-0.2971	-0.2060
Eli Stone	#DIV/0!	-0.0850	0.1196	0.1196	#DIV/0!	0.1934	-0.5436	0.0262	-0.1723	Eli Stone	#DIV/0!	-0.0850	0.1196	0.1196	#DIV/0!	0.1934	-0.5436	0.0262	-0.1723
ER	-0.0708	-0.1195	0.6452	0.1160	0.3861	0.4488	0.0153	0.1456	0.4825	ER	-0.0708	-0.1195	0.6452	0.1160	0.3861	0.4488	0.0153	0.1456	0.4825
Friday Night Lights	0.0605	-0.0672	0.8969	0.3315	-0.2739	0.0366	-0.0347	0.8125	-0.1586	Friday Night Lights	0.0605	-0.0672	0.8969	0.3315	-0.2739	0.0366	-0.0347	0.8125	-0.1586
Gossip Girl	0.6956	0.4608	0.5770	0.3935	0.4312	0.4850	0.3368	0.6711	0.5228	Gossip Girl	0.6956	0.4608	0.5770	0.3935	0.4312	0.4850	0.3368	0.6711	0.5228
Grey's Anatomy	0.1912	0.2137	-0.5101	-0.0939	0.0297	0.0629	0.1450	-0.1281	0.1151	Grey's Anatomy	0.1912	0.2137	-0.5101	-0.0939	0.0297	0.0629	0.1450	-0.1281	0.1151
Hannah Montana	0.2652	0.3080	0.4887	-0.2777	-0.1284	0.0305	0.2034	0.6612	0.0598	Hannah Montana	0.2652	0.3080	0.4887	-0.2777	-0.1284	0.0305	0.2034	0.6612	0.0598
Heroes	-0.3030	-0.2748	0.6889	0.1071	0.2253	0.0230	-0.3773	0.2077	-0.1333	Heroes	-0.3030	-0.2748	0.6889	0.1071	0.2253	0.0230	-0.3773	0.2077	-0.1333
House M.D.	-0.3814	-0.1731	-0.1328	-0.0853	-0.0598	-0.1566	0.1121	-0.3391	-0.0685	House M.D.	-0.3814	-0.1731	-0.1328	-0.0853	-0.0598	-0.1566	0.1121	-0.3391	-0.0685
iCarly	0.6911	-0.0331	-0.3027	0.1428	-0.1157	0.7091	0.3912	0.0906	0.4576	iCarly	0.6911	-0.0331	-0.3027	0.1428	-0.1157	0.7091	0.3912	0.0906	0.4576
Knight Rider	-0.0153	-0.0388	-0.2735	-0.1893	0.2565	0.4911	-0.2124	-0.1686	0.3615	Knight Rider	-0.0153	-0.0388	-0.2735	-0.1893	0.2565	0.4911	-0.2124	-0.1686	0.3615
Law and Order	0.0337	-0.2988	0.8095	0.7465	-0.4490	0.7781	-0.9541	0.0337	-0.1770	Law and Order	0.0337	-0.2988	0.8095	0.7465	-0.4490	0.7781	-0.9541	0.0337	-0.1770
Law and Order: CI	0.0349	-0.0999	0.1862	0.0201	0.0206	0.1858	0.1726	0.1350	0.1877	Law and Order: CI	0.0349	-0.0999	0.1862	0.0201	0.0206	0.1858	0.1726	0.1350	0.1877
Life	-0.0853	-0.0682	0.5754	-0.0476	-0.1640	-0.2809	-0.2124	0.4837	-0.4064	Life	-0.0853	-0.0682	0.5754	-0.0476	-0.1640	-0.2809	-0.2124	0.4837	-0.4064
Life on Mars	-0.5960	#DIV/0!	-0.3865	-0.1587	#DIV/0!	#DIV/0!	-0.1560	-0.5531	-0.2144	Life on Mars	-0.5960	#DIV/0!	-0.3865	-0.1587	#DIV/0!	#DIV/0!	-0.1560	-0.5531	-0.2144
Lost	0.3656	0.1834	-0.4122	0.2842	0.3068	0.2963	-0.2306	0.1623	0.2252	Lost	0.3656	0.1834	-0.4122	0.2842	0.3068	0.2963	-0.2306	0.1623	0.2252
Medium	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.3578	0.3858	#DIV/0!	0.0211	Medium	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-0.3578	0.3858	#DIV/0!	0.0211
My Name is Earl	-0.1986	0.0838	-0.0168	0.0704	-0.3764	0.1264	0.0000	-0.0660	-0.1492	My Name is Earl	-0.1986	0.0838	-0.0168	0.0704	-0.3764	0.1264	0.0000	-0.0660	-0.1492
NCIS	-0.3756	-0.1625	0.3494	-0.1632	-0.1012	0.2157	-0.0050	-0.1014	0.0841	NCIS	-0.3756	-0.1625	0.3494	-0.1632	-0.1012	0.2157	-0.0050	-0.1014	0.0841
Numb3rs	0.5435	0.8066	0.0166	-0.0648	-0.2482	-0.4563	-0.3293	0.6214	-0.5917	Numb3rs	0.5435	0.8066	0.0166	-0.0648	-0.2482	-0.4563	-0.3293	0.6214	-0.5917
Prison Break	-0.0513	-0.2450	-0.3099	-0.1758	0.0445	-0.0973	-0.3115	-0.3526	-0.3455	Prison Break	-0.0513	-0.2450	-0.3099	-0.1758	0.0445	-0.0973	-0.3115	-0.3526	-0.3455
Private Practice	0.9970	0.6515	-0.2334	0.9583	0.3525	0.9717	0.0724	0.9166	0.7276	Private Practice	0.9970	0.6515	-0.2334	0.9583	0.3525	0.9717	0.0724	0.9166	0.7276
Psych	-0.4334	0.1380	-0.2370	-0.1726	0.0574	0.2086	-0.2370	-0.2709	0.0864	Psych	-0.4334	0.1380	-0.2370	-0.1726	0.0574	0.2086	-0.2370	-0.2709	0.0864
Pushing Daisies	-0.3927	-0.2110	0.3497	-0.0462	0.3633	-0.1060	0.1895	-0.0914	0.2336	Pushing Daisies	-0.3927	-0.2110	0.3497	-0.0462	0.3633	-0.1060	0.1895	-0.0914	0.2336
Scrubs	0.0570	-0.0987	-0.0947	0.0554	-0.0812	0.2191	0.3381	-0.0482	0.3970	Scrubs	0.0570	-0.0987	-0.0947	0.0554	-0.0812	0.2191	0.3381	-0.0482	0.3970
Smallville	0.7307	-0.2773	0.1416	0.2433	0.6037	0.2693	-0.3491	0.4724	0.4044	Smallville	0.7307	-0.2773	0.1416	0.2433	0.6037	0.2693	-0.3491	0.4724	0.4044
Terminator: Sarah Con	0.3074	0.1440	0.3020	0.1978	-0.1700	0.1184	-0.1687	0.4561	-0.1478	Terminator: Sarah Con	0.3074	0.1440	0.3020	0.1978	-0.1700	0.1184	-0.1687	0.4561	-0.1478
The Office	0.1014	0.0739	0.0559	-0.2266	-0.0768	-0.3396	-0.1897	0.1229	-0.3359	The Office	0.1014	0.0739	0.0559	-0.2266	-0.0768	-0.3396	-0.1897	0.1229	-0.3359
Ugly Betty	0.0121	0.0627	0.1141	0.2231	0.3415	0.0864	0.1662	0.1312	0.3188	Ugly Betty	0.0121	0.0627	0.1141	0.2231	0.3415	0.0864	0.1662	0.1312	0.3188
Without a Trace										Without a Trace									

When negatives were zeroes out, not all shows saw a stronger correlation. House M.D. and Law and Order: Criminal Intent both retained meaningful correlation, but replacing negative numbers with zeros made the correlation slightly weaker. Heroes went from slightly meaningful to not meaningful. Medium went from meaningful to no meaning because of a divide by zero issue. Life and Psych did not change at all. Terminator: The Sarah Connor Chronicles went from strong meaningful correlation to meaningful correlation.

The three-day period after an episode airs has more shows with meaningful correlations than in the three day period before an episode airs: 15 after and 11 before. When only very meaningful correlations are looked at, the picture changes some: 5 very meaningful correlations before and 2 after. The shows with meaningful correlation after an episode airs are 24, CSI, CSI: Miami, ER, iCarly, Knight Rider, Law and Order: Criminal Intent, Life, Numb3rs, Prison Break, Psych, Scrubs, Smallville, Terminator: The Sarah Connor Chronicles, Ugly Betty and Without a Trace. Of these fifteen shows, only three have negative meaningful correlations. They are Life, Prison Break and Ugly Betty.



Individual days in that three-day period are less predictive than when combined together. There are 7 meaningful correlations on the day after, 10 two days after, and 10 on the third after an episode airs. On the second day, four correlations were negative. On the third day, five correlations were negative. In several cases, these negative correlations on specific days meant that there was random correlation for the whole three-day period. This was the case for Desperate Housewives, Eli Stone, Heroes, Medium and Private Practice.

## Conclusion

Fan fiction communities are well established and producers are very aware of fan production of this content. References to it have begun to be referenced in the canon of shows such as *Crossing Jordan*, *CSI* and *Glee*. The consensus amongst fans, especially those desiring certain plot lines and relationships to appear on the show, is that their material is reflective of wider interest in a show.

The results are rather clear and support this theory: The more fan fiction that is published in the three day period before and after, and on the day an episode airs, the higher the Nielsen Rating should be. The comparatively large volume of stories posted in the three-day period before the airing of the January 15, 2009 episode of *CSI*, one of the highest ratings for any show in this sample, is thus no surprise. The same is true for the poorly rated September 25, 2008 episode of *CSI*, which had one of the lowest three day periods for new stories. This is pretty constant and there are very little variations from this general pattern for genres and television networks except for shows on USA, where their pattern is reversed. On a show by show basis, the patterns can and do change some. *Burn Notice*, *Eli Stone*, *Grey's Anatomy*, *Law and Order: CI*, *Prison Break* have periods where they defy the theory but they are exceptions, not the rule and bigger sample sizes would likely see changes in these results. On the whole, this information is reliable and can be useful for making decisions.

This information can be helpful to producers and others dependent on predicting Nielsen Ratings. Content producers can look at activity levels and determine audience interest and try to use that information for decisions down the road. Advertisers can try to predict the ratings and determine if their ad buy during a show was worthwhile. If correlation exists between performance in the US market and their own, non-American networks can use fan fiction to determine level of promotion of certain story lines before they air an episode. Many of these decisions require more lead time than three days so the usefulness of this data is not necessarily that for easy to identify decisions.

For short-term decisions, this method can be a cheaper and more time effective alternative to monitoring social media volume. Studies done by organizations like Networked Insights's one that was discussed on a Mashable article on October 28, 2009 suggest that social media interactions may have some correlation with Nielsen ratings: The more social media interactions, the more likely the show is to have higher Nielsen Ratings. Fan fiction as a metric is easier and cheaper to quantify than social media interactions.

The real value of this can be used in long-term decision making when they are creating the product because fan fiction can often provide insight into fan thinking. Fans let down their guard and write what they really desire to see happen and share it with the intention of bonding with other fans. The material is not created with the idea of harassing writers and producers to make casting changes, explore certain plot lines, favor certain characters or to make other changes in the show. Fan fiction stands on its own in terms of the community, its content and isolation from producers. This is different than message boards, blogs and other forms of social media. There fans often pitted against each other, have agendas to try to change the minds of production staff, try to get actors to change interpretations of the characters or force a network to make changes to cater to their interest group. In this

context, a textual analysis can be done looking for common elements for fan fiction based around shows on days when they have higher Nielsen Rating and higher levels of fan fiction. The common themes can then be integrated into the show down the road to help try to improve Nielsen Ratings in the long term.

Additional value for this information could be gained for vested interests by expanding upon the initial research by increasing the size of the data set in several areas. These include number of archives in the sample, increase the number of shows to include shows that have started on the air since the data was originally compiled, increase the time period represented, separating sites that cater to certain nationalities and comparing that to the Nielsen equivalents in those countries, and comparing fan fiction posting volume to online viewing volume. This research would improve the overall reliability of the data, and allow for application for other genres not covered in this paper like soap operas, cartoons and anime. It would also help vested parties who are based out of the United States and who are dependent on episodic television that appears in other mediums make more informed decisions. This additional research needs to be done and should help vested parties make better decisions.