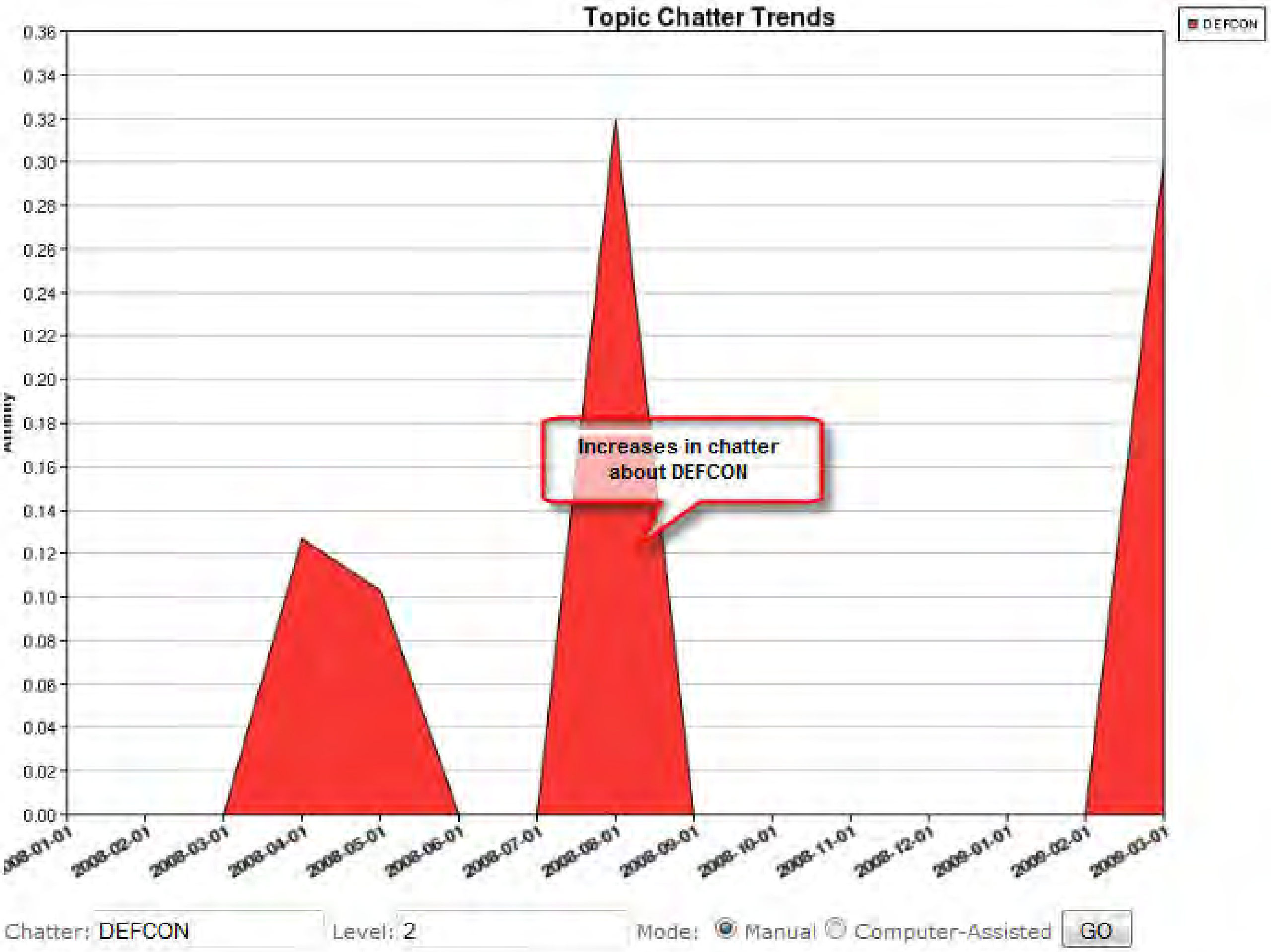
V The Veritas Project

Understanding the Security Trends Interface

So how does this thing work? Basically the premise of this study is to track increases in "chatter" (based on keywords) just like what ECHELON, INTERPOL, or the other intelligence gathering organizations does out there. This is based on information gathered through forums, websites, blogs, and data from various contributors. This how-to presents a quick overview of how this whole thing works.

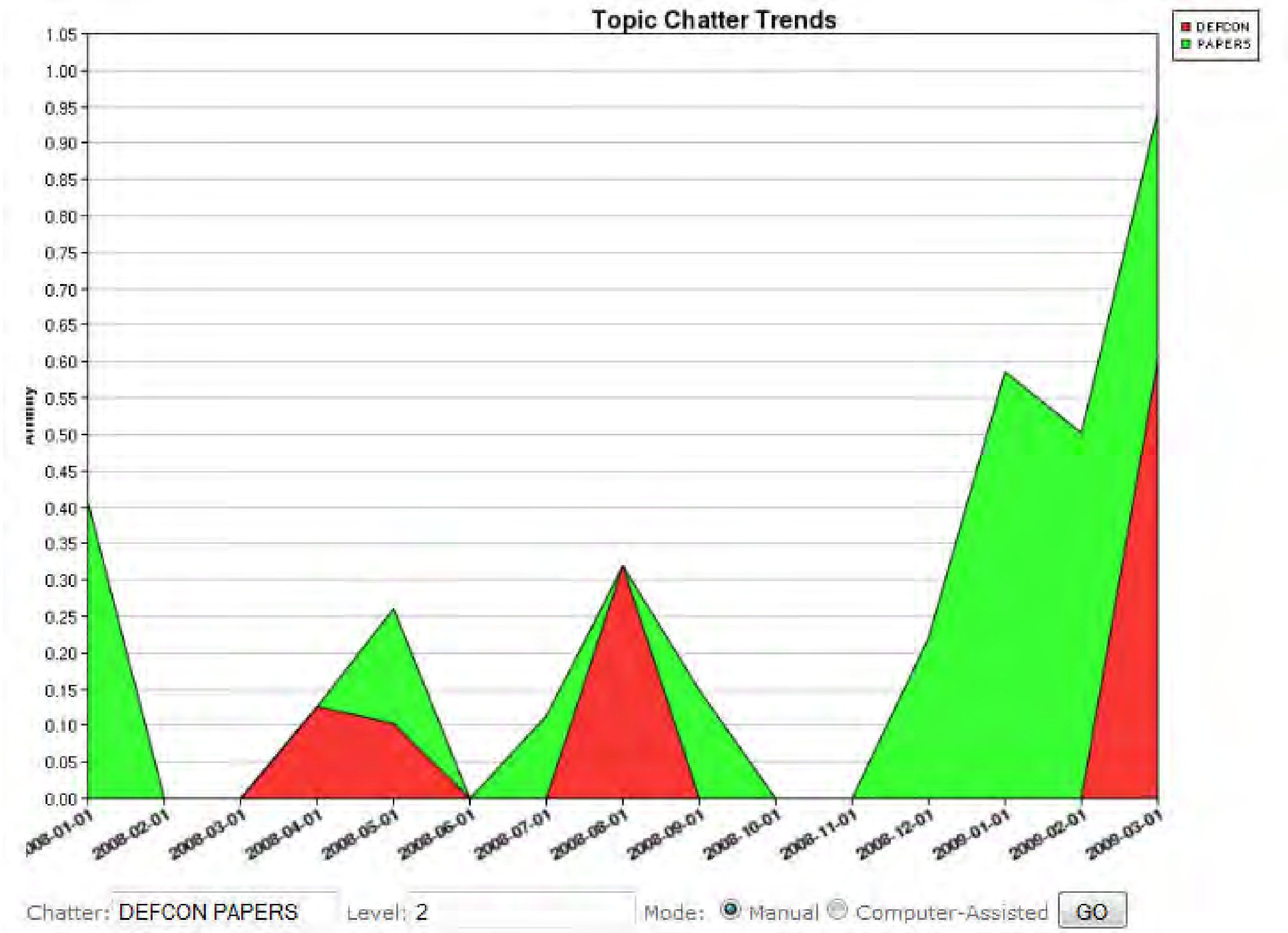
Let's start off with a simple data mining exercise. For this one, let us search for increases in chatter on the keyword "DEFCON". This is done by entering DEFCON in the "chatter" field and putting the mode as "manual". I'll explain later the difference between "manual" and "computer-assisted" later as well as the "level" field. Press "Go" and you will get this:



Increases in chatter about DEFCON

Chatter: DEFCON Level: 2 Mode: ● Manual ○ Computer-Assisted

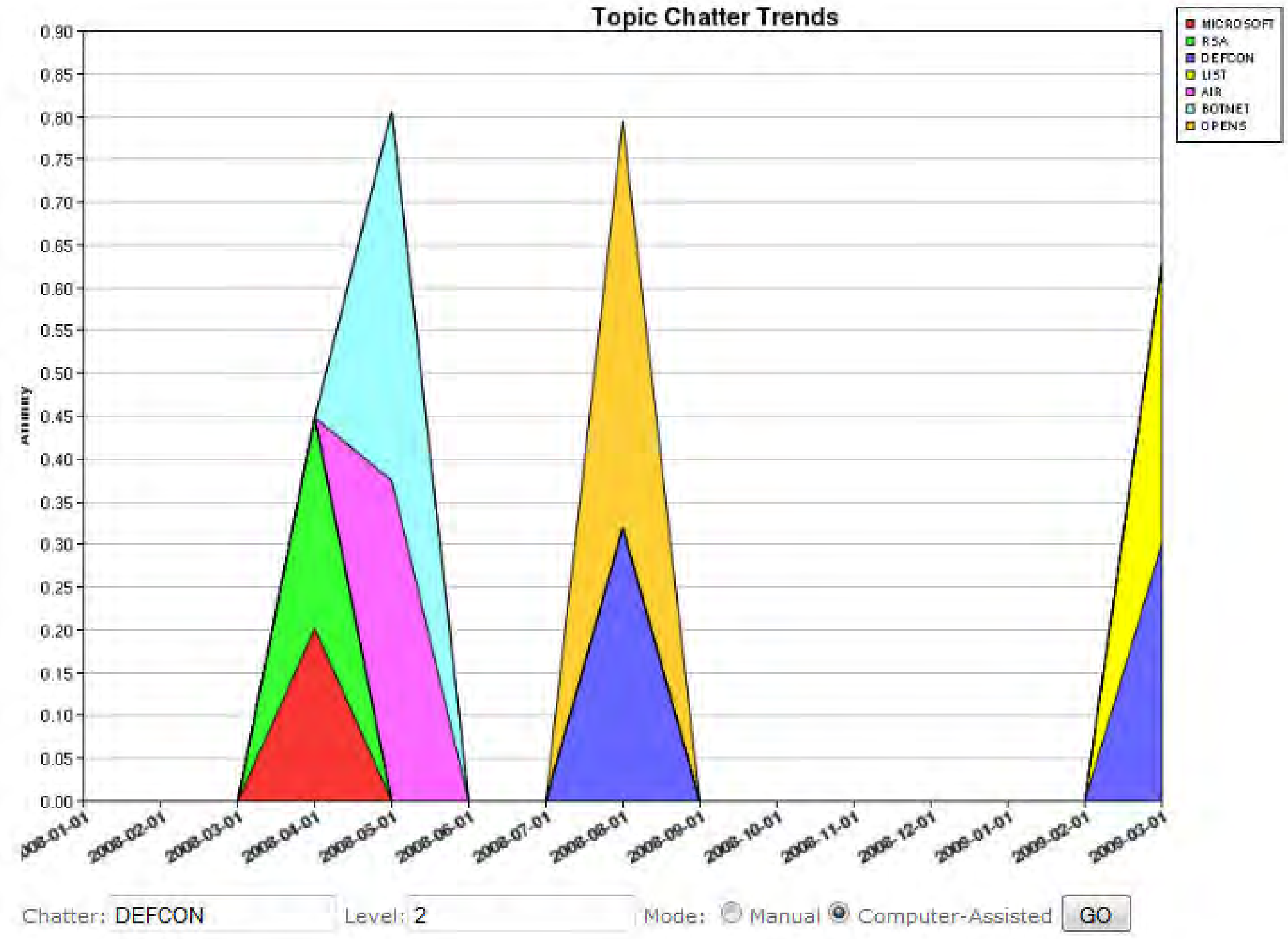
Note that the increases in DEFCON chatter expectedly corresponds around the time for call for papers and the conference itself. Next let's try overlaying it with something we know that is related to DEFCON, the keyword "PAPERS". Using the same instructions, we add the word "PAPERS" in the chatter field separated by a space. Here's what we get:



Chatter: DEFCON PAPERS Level: 2 Mode: ● Manual ○ Computer-Assisted

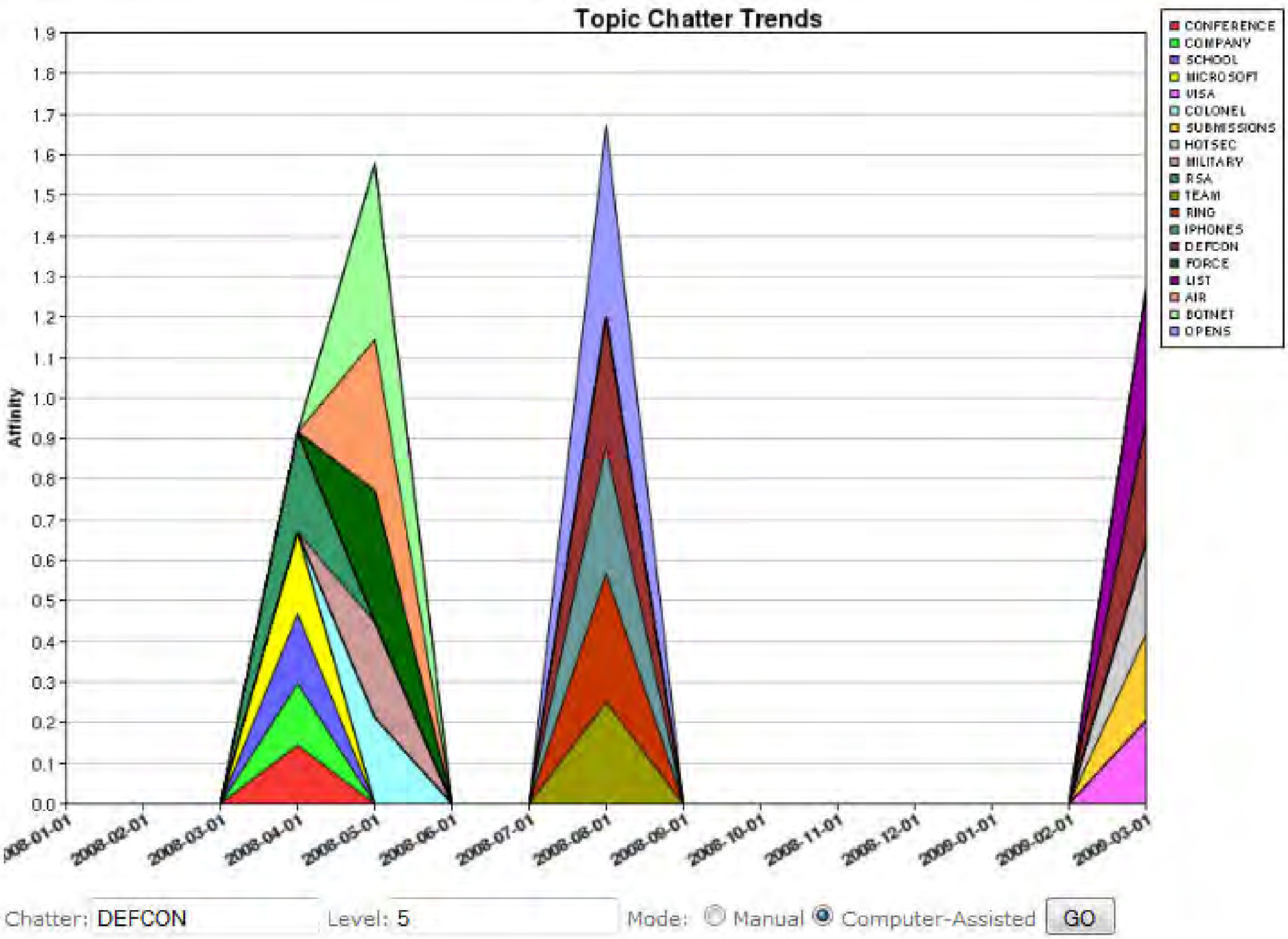
Note that increases in chatter on the keyword "PAPERS" happen before and after the CFP and before and after the conference itself. Notice how closely the the increases in chatter follows. You'll notice a strange increase in "PAPER" chatter around the end of the year. Could this be still related to DEFCON? Possibly, but let's look more closely.

The next demo will show you the "computer-assisted" mode. The computer-assisted mode let's the computer or more specifically, the AI algorithms to determine which keywords are related to the topic we are researching. Remember, in our first search, we manually determined that there could be a relationship between DEFCON and PAPERS. This time, we will let the computer decide for us what keywords are related with each increase in DEFCON chatter. To do this, just type in DEFCON, leave the level to 2 and then check "computer-assisted" as the mode. After pressing "Go" you'll get this:



Chatter: DEFCON Level: 2 Mode: ○ Manual ● Computer-Assisted

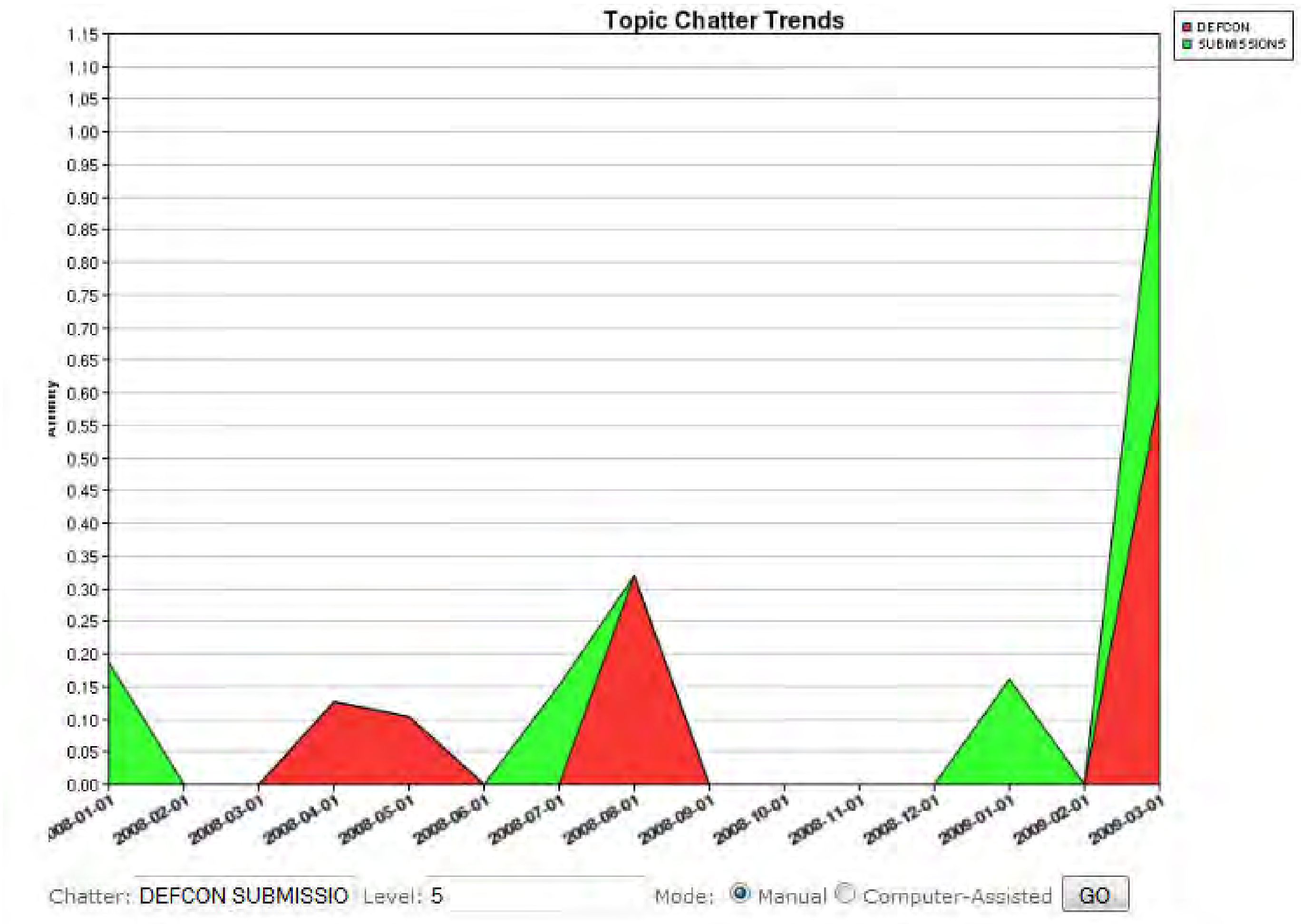
In this result, the algorithms running this process determined that the words on the legend located in the right hand side are those that are the most related to the keyword DEFCON. The increases in chatter for each related word is plotted in the graph. You'll see 7 topics that the algorithm deemed are the most related, if you want to see more related keywords, you can adjust the level by changing the "level" field. Try changing it to 5 and you'll get this:



Chatter: DEFCON Level: 5 Mode: ○ Manual ● Computer-Assisted

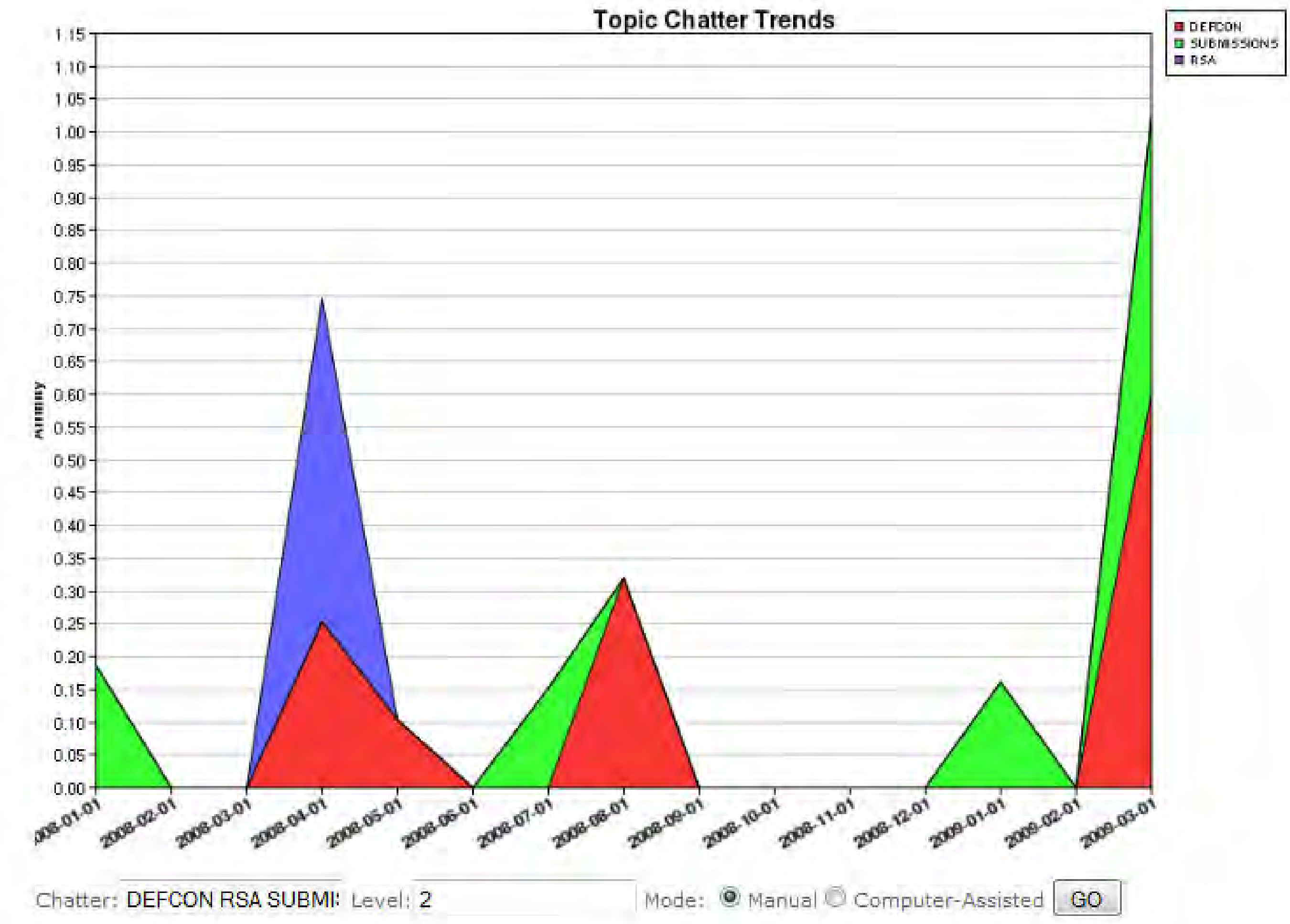
Note that there are more keywords now that the algorithms are identifying as related to DEFCON. The higher the level, the more related keywords the algorithms tries to plot in the graph. Sometimes, this could be a bit overwhelming and due to the nature of artificial intelligence algorithms, sometimes not entirely accurate (think Terminator and I,Robot - misinterpreting stuff like killing us all to save the earth), so you'll have to fine tune your searches and temper it with good old human intuition and interpretation.

You'll notice that one of the related keywords that the algorithms are identifying is actually "SUBMISSIONS". To see a clearer view on how it relates to the keyword DEFCON, we can change the mode back to "manual" and overlay it with the "DEFCON" like what we did in "PAPERS". Here's what you'll get:



Chatter: DEFCON SUBMISSIO Level: 5 Mode: ● Manual ○ Computer-Assisted

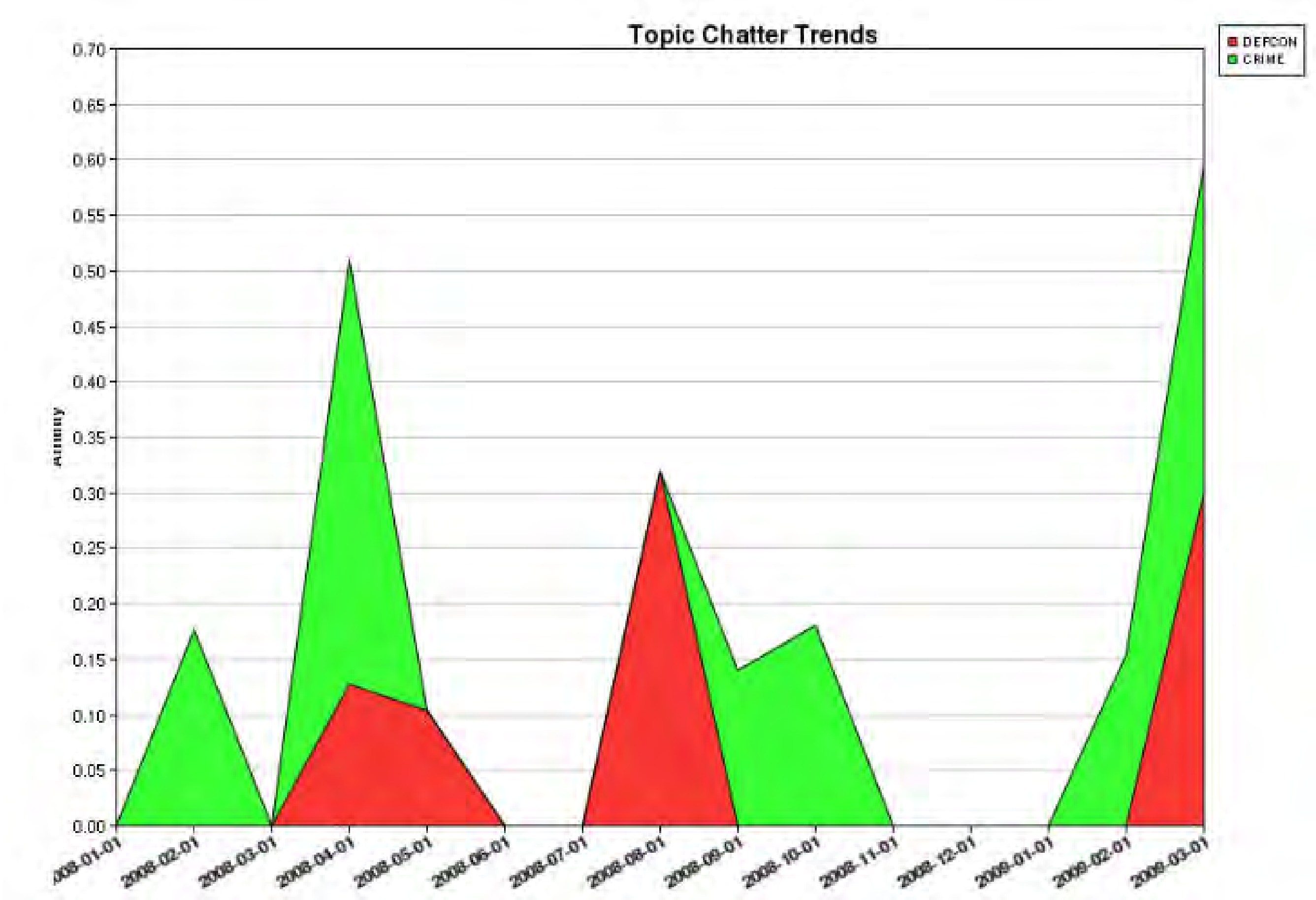
Note that it is fairly similar to the "PAPERS" search except that the "SUBMISSION" increases in chatter happen after the CFP and before the actual conference which makes sense since submissions are done before the conference starts unlike papers which could have increases in chatter before and after the conference. Notice that we still see some unusual spikes in submission at the end and beginning of the year which may or may not be totally related to DEFCON. One explanation of this spike could be seen in the "computer-assisted" search that we did before. Note that another security conference appears there which is "RSA". Let's try plotting "DEFCON", "RSA", and "SUBMISSIONS" in manual mode:



Chatter: DEFCON RSA SUBMI: Level: 2 Mode: ● Manual ○ Computer-Assisted

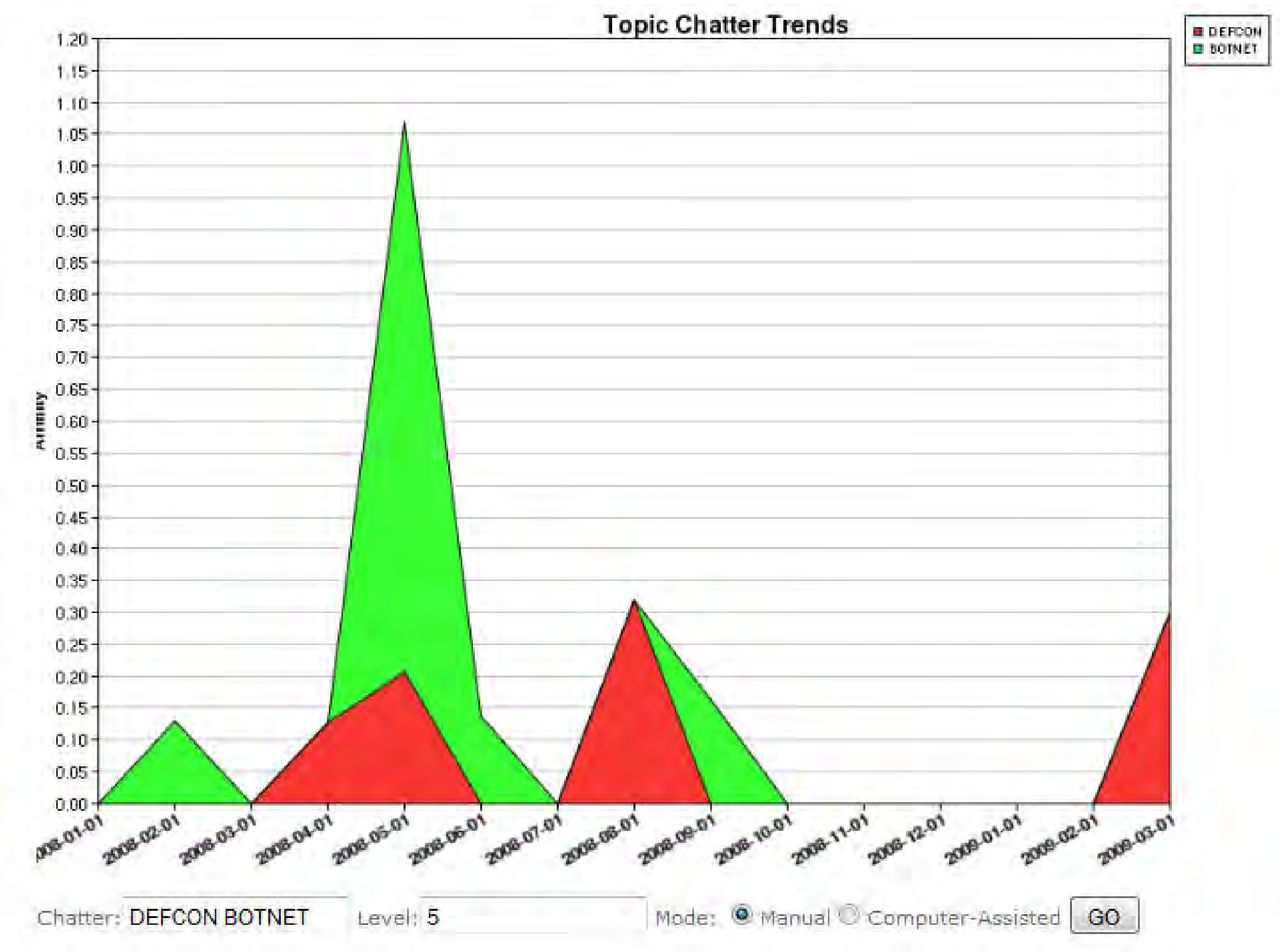
Note that in this search you'll see an increase in chatter of the "RSA" keyword around April. This could be a possible hypothesis explaining the increase in chatter on "SUBMISSIONS" on the period before that. Of course, there could be other explanations if you do other correlations but for this view, it does make sense.

We can always say that surveillance, threat intelligence and similar activities like this one can be more art than science and depends a lot on the analyst that does the interpretation of the chart. For example, let's do a search on "DEFCON" and "CRIME":



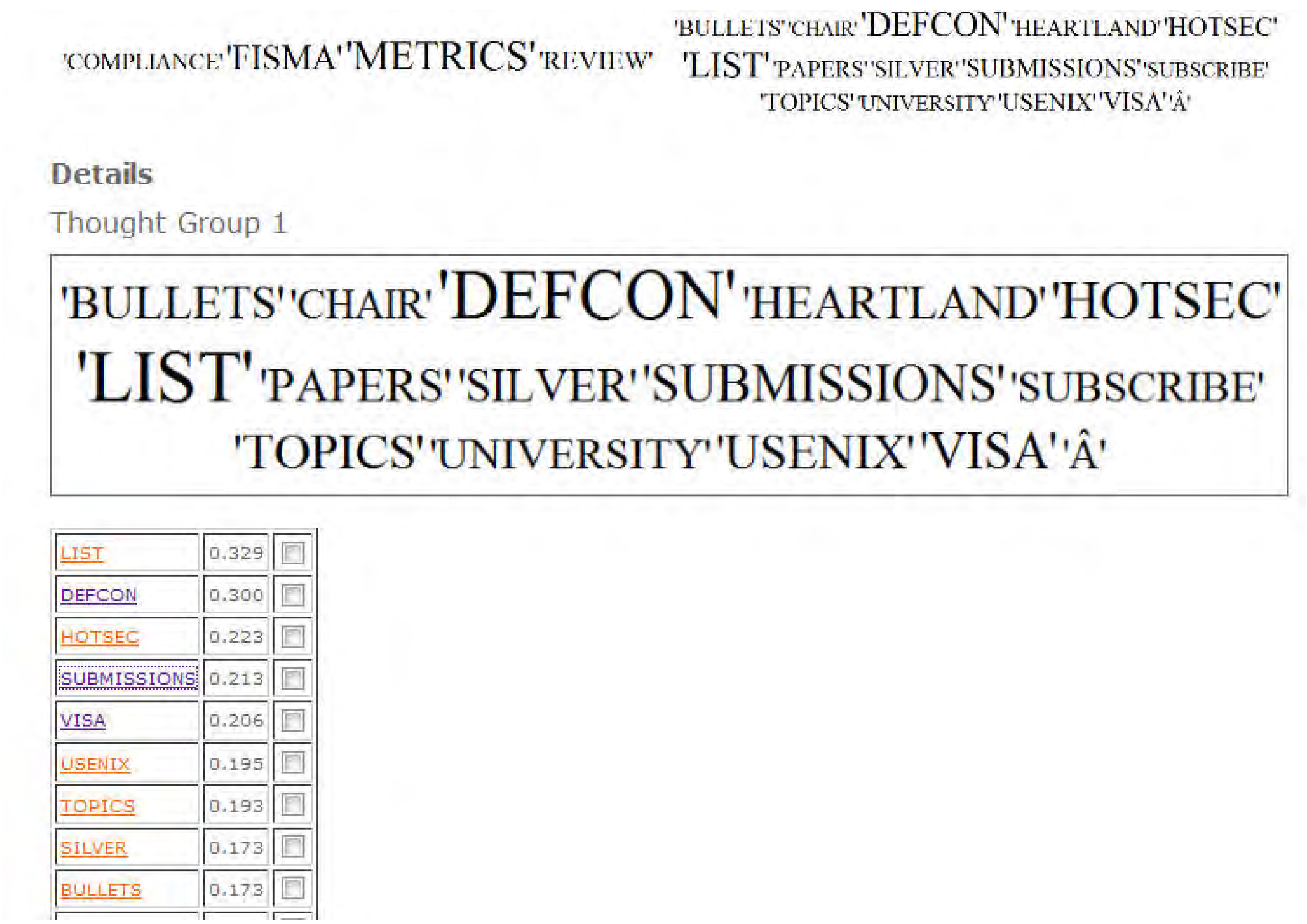
Note that after the actual conference there is an increase in chatter on the keyword "CRIME". Obviously, it will be irresponsible to say that the confernce promotes an increase in crime right? If I was the Central Terminator AI, then DEFCON won't be here anymore. Thus it is safe to say that before you make conclusions, remember to do more in depth "human" research. Remember that this tool is meant to assist you in identifying trends, not dictate what the trends are. Looking at the graph though, it is certainly an interesting coincidence that is worth looking at.

Here's another example, this time an overlay between "DEFCON" and "BOTNET". The keyword "BOTNET" actually shows up using the computer-assisted mode which means that it has a high correlation with the keyword "DEFCON":



Chatter: DEFCON BOTNET Level: 5 Mode: ● Manual ○ Computer-Assisted

Note that there's a very noticeable increase in chatter on the keyword "BOTNET" around the times that "DEFCON" chatter increases also. Aside from, this you can also go to a monthly view which shows "clusters" which are basically words that are related to each other. This example shows the March 2009 Monthly Topic Analysis:



Details

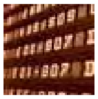
Thought Group 1

LIST	0.329	□
DEFCON	0.300	□
HOTSEC	0.223	□
SUBMISSIONS	0.213	□
VISA	0.206	□
USENIX	0.195	□
TOPICS	0.193	□
SILVER	0.173	□
BULLETS	0.173	□

Notice that group of words has probably has a high correlation with the DEFCON call for papers. Not too sure why Heartland appears there too but there's probably a link that will come up with further research. Note that the words listed below the cluster can be clicked and this will automatically be plotted in the graph in manual mode so you can see the increases in chatter for that keyword throughout the time period available. A facility to cross-correlate and search in Google is also provided to provide the "human" research factor that I have been mentioning throughout this how-to.

Hope my explanation made sense. Happy hunting!

Click here to go back to the study



Suggest a Topic
Want us to data mine a topic?
Send us your topic at veritasproject@gmail.com

Do It Yourself
Want to learn to make these?
Jumpstart your knowledge with these resources. Go!

Contact Us
Email: veritasproject@gmail.com

top