
1. How does Signals define a Shift?

A shift is a specific time frame with a notable increase or decrease in the volume of mentions related to a topic within the time period specified by a user's filters.

The purpose of shift detection is to alert users to a sudden increase or decrease of these variables so that they can investigate *why* there is such an increase/decrease.

The core methodology is to smooth the data with a Savitzky–Golay filter, which removes spikes & seasonality from a time series (e.g. volume by days, interactions by weeks, etc.) to get the true “trend line.”

Every time there's a change of direction, we get a time frame. All time frames are then scored with:

$$\text{score} = (\text{end_value} - \text{start_value}) / (\text{length})$$

More information about the Savitzky–Golay filter can be found [here](#) and [here](#).

2. How does Signals define a Peak?

Peaks reflect the maximum value of any growing trend and point to the most significant dates related to good or bad buzz. Identifying these moments without Signals is difficult considering dashboards contain high-volume and low-volume topics.

The basic methodology for peak detection is to take the first order difference of a univariate series and to identify peaks as shards with values higher than the local threshold:

$$\text{thres} = \text{thres} * (\max(y) - \min(y)) + \min(y)$$

To score the identified peaks, the following methodology is applied:

- Sort all detected peaks by standard deviation
- Normalize results for each metric

- When n cards have the same score, then sort them by the already computed standard deviation
-

3. How does Signals define a Correlation?

The correlation algorithm in Signals aims to identify topics that tend to evolve in the same way (strong positive correlation), or in the opposite way (strong negative correlation).

The basic methodology for correlation detection is as follows:

For each topic:

1. Apply a time series interval condition, based on user defined filters.
2. Define the time series by **days** when the time period is < 92 days. If the time period is >= 92 days & < 366 days, define time series by **weeks**. If the time period is >= 366 days, define the time series by **months**.
3. Compute correlation algorithm with all of the other topics

The algorithm applies the Spearman correlation coefficient and the first order difference of the time series.

More information about the Spearman correlation coefficient can be found [here](#) and [here](#).

4. How does Signals define a Co-Occurrence?

A high co-occurrence means that the topics in question tend to be mentioned at the same time (within one mention).

The basic methodology for identifying co-occurrences is to take the intersection value between two topics, with one of those topics designated as a reference. The co-occurrence score is then calculated as follows for topic A and topic B, with topic A designated as the reference:

$$\text{score} = (A \cap B) / A$$

5. How does Signals define a Daily Pattern?

A daily pattern identifies which weekday(s) a specific topic tends to be mentioned most often. The daily pattern Signal is the first of many steps to come in providing users with an on-demand "media kit" containing all of the information relevant to how brands and consumers tend to post about topics of interest.

The methodology for identifying daily patterns of interest is to:

1. Determine the total volume of mentions, interactions or impressions for a topic over the entire time period specified by the user
 2. Determine the total volume of mentions, interactions, or impressions for the topic by each weekday
 3. Compute the ratio between the totals for each weekday and the total volume over the entire time period
-

6. What is Synthesio's definition of Viral (for Media Signals)?

Measuring how viral an image or video is through the computation of the interactions it has generated. As of now, we don't provide the ability to identify "trending" videos, articles or media; because we do not store the evolution of interactions (aka interactions time series).

In the Signals module, a piece of media is said to be viral if its publisher has a large follower base and manages to generate a lot of engagement (likes, replies, etc.).

The "virality" of a piece of media is calculated as follows:

virality = influence * interactions

Whereby ...

- Influence = number of followers at the time of posting
 - Interactions = total number of likes, replies, etc.
-

7. How does Signals define top words in the Context page?

The Context page is designed to push top words aiming to explain why topics were mentioned online. Therefore, we exclude words or hashtags from data queries (from setup).

We provide 5 different ways to help users visualize these top-words so understand context behind topics mentions:

- Top-50 words in general, from all publishers categories mixed, with a network map
- Top-50 words from news media articles

- Top-50 hashtags from social networks (Twitter, Instagram, Facebook)
- Top-50 words from consumers reviews (Amazon, Tripadvisor...)
- Top-50 words from forums (including Reddit)

Top words are found by tokenizing mentions, removing stop words & words from queries, then counting frequency for each token (word).

The Network map displays top-50 from the whole mentions corpus, with data filters applied

- Dots are words
- Links are co-mentioned words intensity
- Colors & sizes reflect word frequency

When hovering a dot, you can overview the most co-mentioned words

8. How are Top Signals ranked?

The Signals appearing in the Top Signals page are not ranked. We only display the most significant Signals, from each category, with a few variations (best from peaks in volume, best from peaks in interactions, etc.)

9. How are Signals within each category ranked?

The Signals in each category are ranked by their scores (see questions 1-6).

To summarize, the scores for each category reflect the following:

- **Media:** Large influence and large number of interactions
 - **Peaks:** Big difference between standard deviation and peak value
 - **Trends:** Big difference between start/end values and short duration
 - **Co-occurrences:** Large number of mentions containing two topics at the same time
 - **Correlations:** Similar behavior in terms of mention or engagement volume
 - **Daily patterns:** More mentions on a given weekday
-

10. How is it determined which widgets make up the contextual report?

Widgets tend to provide more information that could be useful to understand why a statistical event has happened. If there's a peak in mentions, one might ask if this is a positive or negative peak, where it comes from, etc.

The widgets available in the current iteration of the contextual reports are those we felt went furthest in answering the "why," "how," and "where" of each Signal. Next iterations will include:

- Words (top and trending)
 - More analytics related to co-occurrences & weekdays
 - More comparisons between statistical event period and global period
-

11. What is the difference between a peak taking place over a week and a shift taking place over a week?

When users define the time period they want to generate Signals, a generic time interval is chosen, depending on the length of this time period:

- Define the time series by **days** when the time period is < 92 days
- Define the time series by **weeks** when the time period is >= 92 days & < 366 days
- Define the time series by **months** when **the** time period is >= 366 days

Shifts must be computed on at least three time series buckets, while peaks are focused on one single bucket.

For example, if a user defines the time period as the last 6 months, then *weeks* are chosen to be the time interval for all analytics.

Additionally, peaks focus on interactions, while shifts reflect mentions volumes for a particular topic.

12. What should I do if I see a Signal that doesn't make sense, or I disagree with?

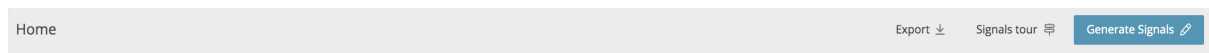
Signals are not based on any subjective factors or variables, but on statistical events; however, if you see a Signal that does not make sense, please let us know by sending an email to signals-team@synthesio.com.

13. Can I customize Signal to look at specific topics or media types?

Yes, that's the purpose of the Signals wizard found under "Update Feed." For more guidance on configuring Signals, check out our [Signals User Guide](#).

14. Can the data found in Signals be exported?

Yes, each Signals page can be exported via the export button found in the top right corner of every page.



15. How does Signals calculate average growth rate for the Trends Landscape Matrix?

The average growth rate is computed by analyzing sub time frames volume growth rates, then computing an average of all these growth rates.

Sub time frames can be days, weeks or even months, depending on the signals study time frame.

Why compute the average?

Because if one topic tends to increase more and more over time, then its volume average growth rate will be more significant than a topic with one huge spike that took place one week, and was then followed by no growth.

If you have any other feedback or questions about Signals, please let us know at signals.team@synthesio.com. And don't forget to look through our [Signals User](#) Guide for detailed instructions about the Signals module.

Helping you discover new insights, and discover them faster, is what gets us excited. We look forward to fine-tuning Signals with you in the coming months.