



**10 OBSERVATIONS ON THE  
ESG DATA  
CHALLENGE**



RavenPack



RavenPack

Forming a full 360-degree view of a company's sustainability profile is a genuine challenge. Most relevant data comes in unstructured formats, which are difficult to consume systematically.

Yet investors are increasingly relying on data-driven approaches to achieve scale and efficiency across ESG use cases. In this report we review 10 observations that help firms address the ESG Data challenge.

**1** *ESG INVESTING IS GROWING BECAUSE IT DELIVERS RETURNS*

**2** *ESG INCREASINGLY RELIES ON ALTERNATIVE DATA & AI ALGOS*

**3** *SENTIMENT EMERGES AS A STRONG OVERLAY SIGNAL FOR ESG*

**4** *USE RISK MOSAICS TO IDENTIFY EMERGING TOPICS*

**5** *ACCOUNT FOR THE MATERIALITY OF NETWORK INFLUENCERS*

**6** *STUDY CO-MENTION NETWORKS TO UNDERSTAND ESG RISK FLOWS*

**7** *CATCH ESG EVENTS WITH A TARGETED TAXONOMY*

**8** *LEVERAGE FACT-CHECKERS TO AUGMENT EXCLUSION LISTS*

**9** *ANTICIPATE MARKET REACTIONS TO ESG ADVERSE MEDIA*

**10** *ELEVATE FROM ESG KEYWORDS TO SUSTAINABLE THEMES*

# 1 ESG INVESTING IS GROWING BECAUSE IT DELIVERS RETURNS

For several decades, the investment community operated under the assumption that restricting investment to companies committed to positive societal outcomes eroded investment returns. Recent studies, however, point to the opposite effect: as ESG criteria become better defined, higher-rated ESG stocks exhibit less severe drawdowns during bear markets. In effect, their inclusion in portfolios provides greater downside protection to investors, which makes more investment sense.

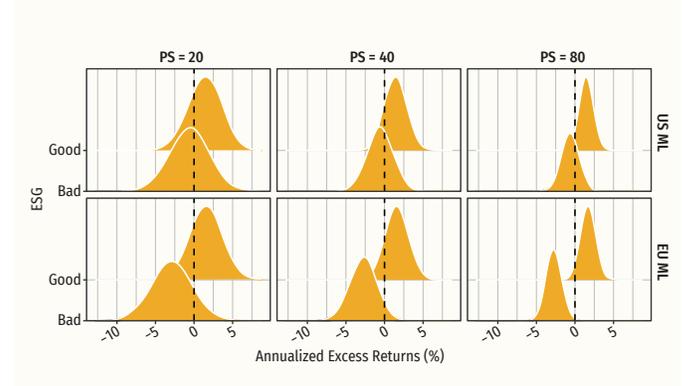
In a recent case study, RavenPack tested the claim made by MSCI, a leading ESG-ratings provider, that higher-rated companies tended to outperform their lower-rated peers on a standalone basis. To examine the impact of ratings, RavenPack created performance distributions by randomly selecting entities having either good or bad ESG ratings according to MSCI.

“Good ESG” distributions outperformed the market across mid- and large-cap stocks in Europe and the U.S., while “Bad ESG” distributions underperformed on a relative basis.

The results were consistent across different portfolio sizes (PS = 20, 40, or 80). Expanding the number of securities resulted in progressively narrower distributions due to greater diversification. The likelihood that a greater proportion of high or low-performance entities would be selected in a random portfolio was higher for smaller portfolio sizes, leading to the longer left and right tails in those cases. It was also apparent that the performance differential was less pronounced in the U.S. relative to Europe, where ESG mandates have been adopted more widely.

The results were qualitatively in agreement with the MSCI findings and support the hypothesis that ESG ratings positively contribute to performance. Results were also consistent across companies with smaller market capitalizations.

▼ Annualized excess return distributions for “good” vs. “bad” ESG companies according to MSCI’s ESG ratings. Source: RavenPack, June 2021



# 2 **ESG INCREASINGLY RELIES ON ALTERNATIVE DATA & AI ALGOS**

ESG investors need to evaluate companies' profile to decide whether to include them in, or exclude them from, their portfolio. Since most ESG ratings are based on self-reported information that's updated infrequently, they can be slow to capture changing perceptions when new data becomes available, and the investment community increasingly considers them insufficient as a stand-alone solution.

With the growing availability of digital information, alternative datasets have emerged to complement these ratings using a more quantitative and systematic approach which can reduce costs, promote more accurate ratings, and raise the prospects of higher alpha.

ESG-relevant signals, from factory incidents to corruption or child labor, are often found in news reports, transcripts, or filings, so sourcing them involves leveraging technology that turns unstructured, textual data into structured insights that can power quantitative ESG approaches. Overcoming that necessary technological hurdle requires considerable storage and computing capabilities usually found only in highly-specialized cloud-based solutions.

Making sense of these data however, from identifying trends to connecting parent companies and subsidiaries, translates into more complexity in investment models. This leads investors to tap into artificial intelligence and machine learning algorithms to extract and clean ESG data, compute indicators and ratings, analyze signals, optimize portfolios and trigger trade execution.

## **A Growing Range of Alternative Data**

Higher volumes and better consistency can be obtained by venturing into unstructured data. Using natural language processing (NLP), for instance, investors can assess whether company management is being honest in its public communications. Monitoring employee satisfaction is another example of using Big Data as an alternative means of measuring a company's social performance. While companies often conduct internal surveys or use HR applications to monitor employee satisfaction, these assessments are unlikely to be accessible to outsiders. Company reviews and ratings by employees at external sources such as Glassdoor.com can provide a similar metric.

# 3 SENTIMENT EMERGES AS A STRONG OVERLAY SIGNAL FOR ESG

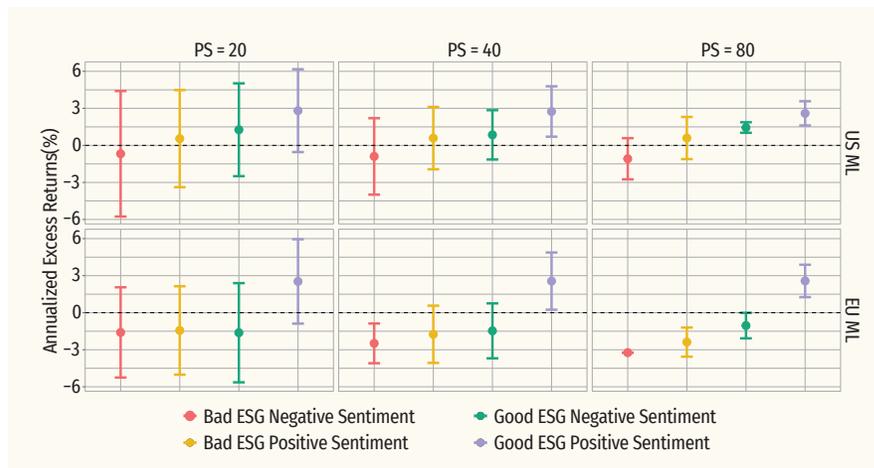
Among indicators derived from alternative data sets, sentiment has emerged as a powerful overlay signal for ESG portfolio returns.

RavenPack sampled stocks from their respective ESG classifications and randomly selected entities with probabilities proportional to their sentiment indicator, creating sentiment-augmented performance distributions.

The distributions progressively narrowed with increasing portfolio sizes due to greater diversification benefits stemming from a lower likelihood of randomly selecting clusters of outperformers or underperformers. The results also showed a more pronounced performance differential in favor of the European universe relative to the U.S.

Positive sentiment was shown to add value to the “Good ESG” clusters, while negative sentiment benefits the “Bad ESG” short strategies. This effect is more apparent when applied to larger portfolios, where the tails of the annualized return distributions exhibit apparent shifts in the direction of sentiment. Similar results were found when evaluating the respective distributions of strategy information ratios.

▼ Annualized excess return distributions for “good” vs. “bad” ESG companies according to MSCI’s ESG ratings, including a “positive” vs. “negative” sentiment overlay Source: RavenPack, June 2021



# 4 USE RISK MOSAICS TO IDENTIFY EMERGING TOPICS

A significant development in the investment industry with applications to ESG is the centralization of multiple datasets into multidimensional risk mosaics. These mosaics can become rich and complex information hubs that allow financial institutions to build a “source of truth” that can be leveraged across the entire organization to democratize content internally.

Risk mosaics can help identify emerging topics and narratives that may show up as abnormalities in the data. Such topics may proceed to grow into major trends driving subsequent market risk across entire sectors or portfolios. Examples include the radical policy shift in the U.S. on climate change after the election of the Biden administration, the emergence of the “Black-Lives Matter” movement, or capturing the first, second, or third waves of the “MeToo campaign.” Beyond fixed event taxonomies, directly tapping the underlying textual content provides an additional layer of flexibility, enabling an internal risk team to capture more nuanced themes and to identify emerging topics that should be on their radar.

Risk mosaic methodologies can also be helpful for quickly evaluating what actions should be taken to protect investments when ESG issues arise. Examples include real-time controversies like scandals, corruption or harassment charges, general litigation proceedings, and natural or human-made disasters. Detecting events using standardized taxonomies allows for creating highly actionable, real-time alerts that can be easily backtested. Moreover, the implementation of automated rules-based decision-making is likely to be more robust using this approach than with standard, unsupervised topic modeling techniques.

When new market themes arise, investors and analysts need to have the ability to implement on-the-fly curation of event topics to track their evolution over time. Using Natural Language Processing techniques, it is possible to monitor a threat over its lifecycle when it first emerges as an unknown-unknown and subsequently develops into a broadly accepted theme that drives risk narratives. For example, users can define a set of sentences that capture a particular topic and translate them into numerical embedding vectors, such that future instances of that topic can be systematically detected based on semantic similarity.

Over the past few years, we have seen several market-driving themes emerge. Brexit, Trump’s trade wars with China, and the Covid-19 pandemic in 2020 have all attracted significant attention from our clients. To develop data mosaic workflows, many organizations have resorted to partnering with leading data vendors and technology providers to develop large-scale data solutions and unlock the full value of content across their organization.

▼ **The Risk Mosaic NLP life cycle** Source: RavenPack, June 2021



# 5 ACCOUNT FOR THE MATERIALITY OF NETWORK INFLUENCERS

What ESG investors look for (their Key Risk Indicators, or KRIs) determines the type of alternative datasets required to compute them. These investors often have to balance both financial and non-financial concerns when constructing portfolios. The maturity of regulation and lack of standardization are added challenges giving rise to the uncertainty that may impact portfolio returns as industry standards evolve. These non-financial risks, however, extend beyond regulation.

The focus on particular topics from special interest organizations (NGOs), the media, or the general public via social media can also impact future regulation and the company's perceived sustainability profile. This dynamic can have a material, real-time impact on businesses. As a result,

companies and investors are increasingly interested in tracking ESG topic lifecycles to manage such risks, further emphasizing the concept of emerging topics. Different sources may affect materiality and market reaction times in diverse ways.

For instance, an emerging topic reaching a critical threshold on traditional or social media may impact a business more acutely than those primarily seeing attention from NGOs or trade and scientific journals. Similarly, any scientific discovery related to environmental research or product innovation can influence regulatory changes and longer-term financial returns.

It is essential to consider this materiality from an event taxonomy perspective and in the selection of sources when building out an ESG data strategy.



▲ ESG Topic Life Cycle Source: RavenPack, June 2021. Logos are trademarks of their respective owners.

# 6 STUDY CO-MENTION NETWORKS TO UNDERSTAND ESG RISK FLOWS

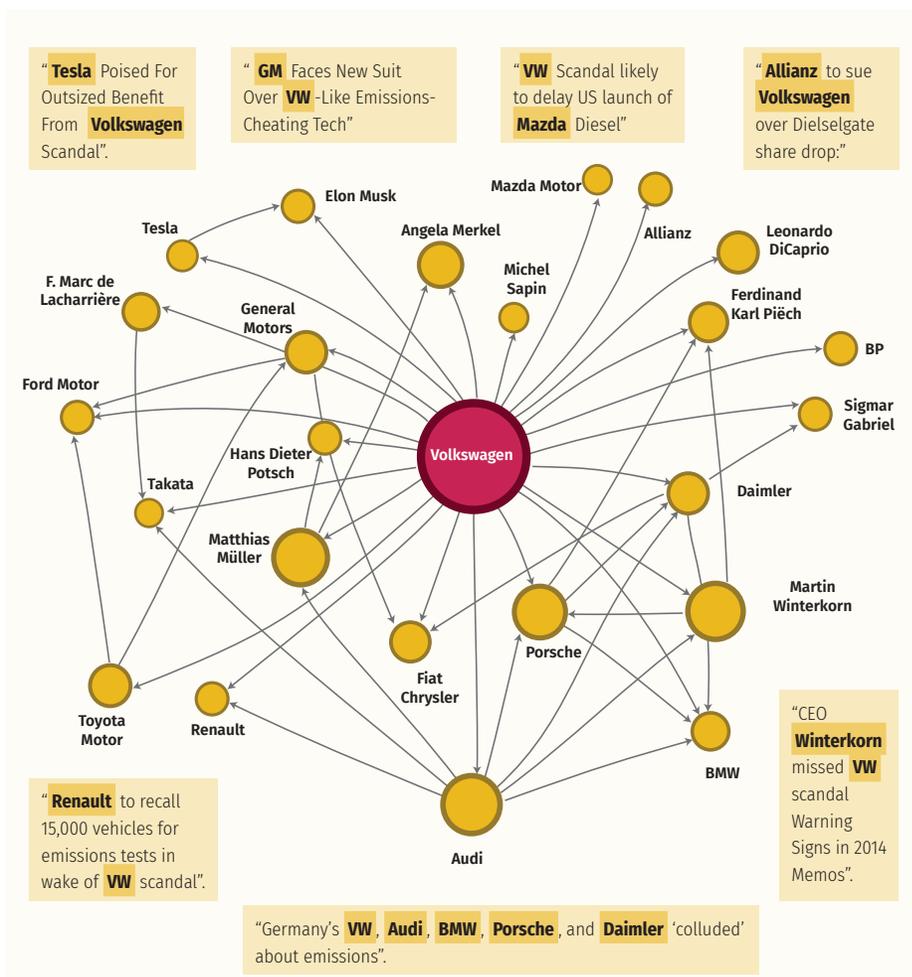
ESG investors can leverage keyword co-mentions in news datasets to build complex networks of entities that share interrelated risks. This concept is best described by example.

In September 2015, the U.S. Environmental Protection Agency (EPA) found that Volkswagen had violated the Clean Air Act by equipping over 590,000 diesel motor vehicles with “defeat devices” designed to circumvent federal emissions tests.

This chart shows the network of entities affected by what was dubbed “Dieselgate”. While Volkswagen was at the center of the scandal, it became apparent that other auto companies were also impacted.

Renault, for instance, was also accused of using emissions-cheating software, while Tesla’s shares rose after the scandal broke as retail investors perceived the electric carmaker as one of the few remaining environmentally friendly auto brands.

The associated network also includes key executives like Martin Winterkorn, then CEO of Volkswagen AG, who allegedly missed internal warning signs that appeared in 2014 memos. Hollywood actor Leonardo DiCaprio was also captured in the network as part of media rumors that he may produce a film based on the scandal.



▲ “Dieselgate” co-mentions Source: RavenPack, June 2021

# 7

# CATCH ESG EVENTS WITH A TARGETED TAXONOMY

News data are exceptionally rich content relative to other alternative data sources it can be consumed in real-time to nowcast traditional financial and economic indicators so ESG investors can be alerted when a new source of risk emerges.

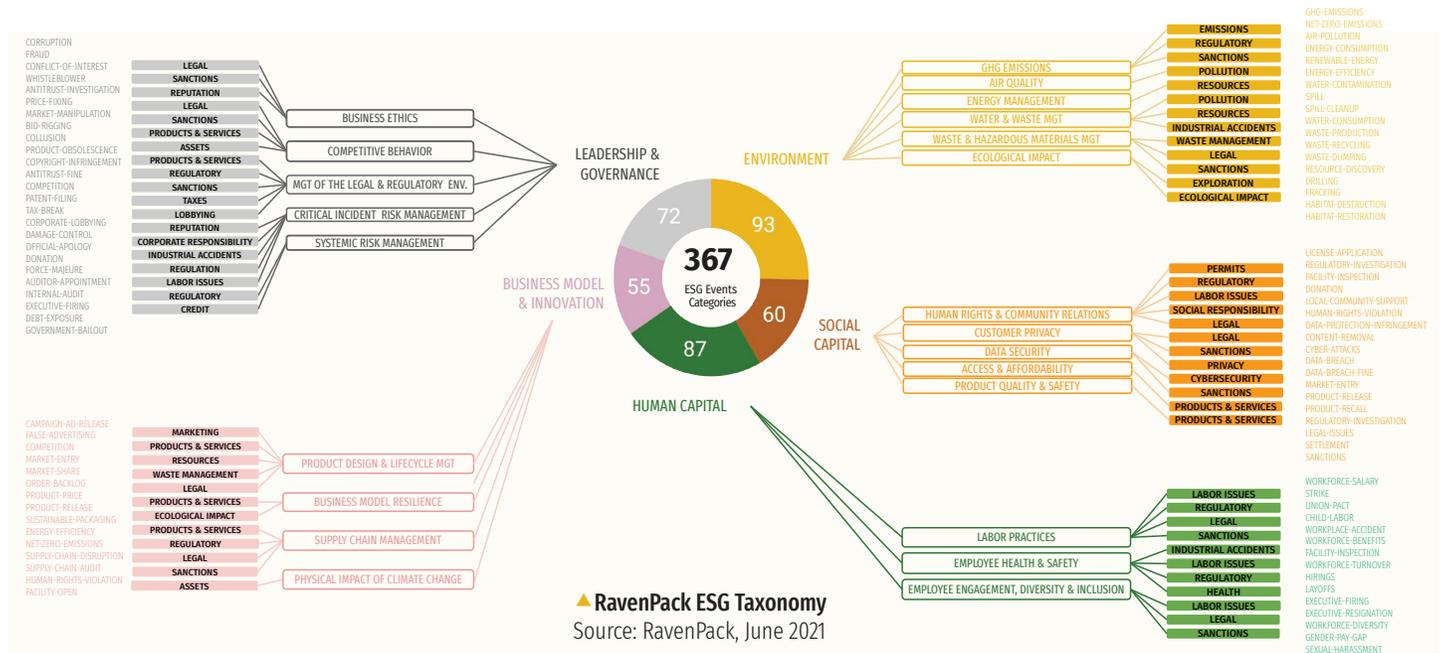
The data benefits from local and global newswire coverage. It covers numerous topics, including business, economic, political, societal, and environmental events – across companies and global macroeconomic entities.

The availability of decades-long news archives provides enough data for comprehensive backtesting and training, allowing users to learn from past events.

To generate an ESG signal from news data, investors rely on a taxonomy. RavenPack has expanded and mapped

its existing, general-purpose event classifications into a specific Ravenpack ESG Taxonomy to address the needs of sustainable investment clients, covering events across Leadership & Governance, Business Model & Innovation, Human Capital, Social Capital, and the Environment.

In total, 367 granular ESG event categories are captured. The Environment branch of the taxonomy, for instance, allows users to track companies that appear in the news linked to greenhouse-gas emissions by mapping their goals and aspirations toward achieving net-zero emissions. Likewise, the Human Capital branch helps oversee any incoming information related to workforce diversity and the gender pay gap, another closely followed set of ESG issues.



# 8

# LEVERAGE FACT-CHECKERS TO AUGMENT EXCLUSION LISTS

ESG investors often use exclusion lists for companies involved in controversial weapons, alcohol, or tobacco. Maintaining these lists often requires a significant amount of manual research. However, tracking news co-mention networks offers a systematic way of expanding the inventory of potential exclusion candidates.

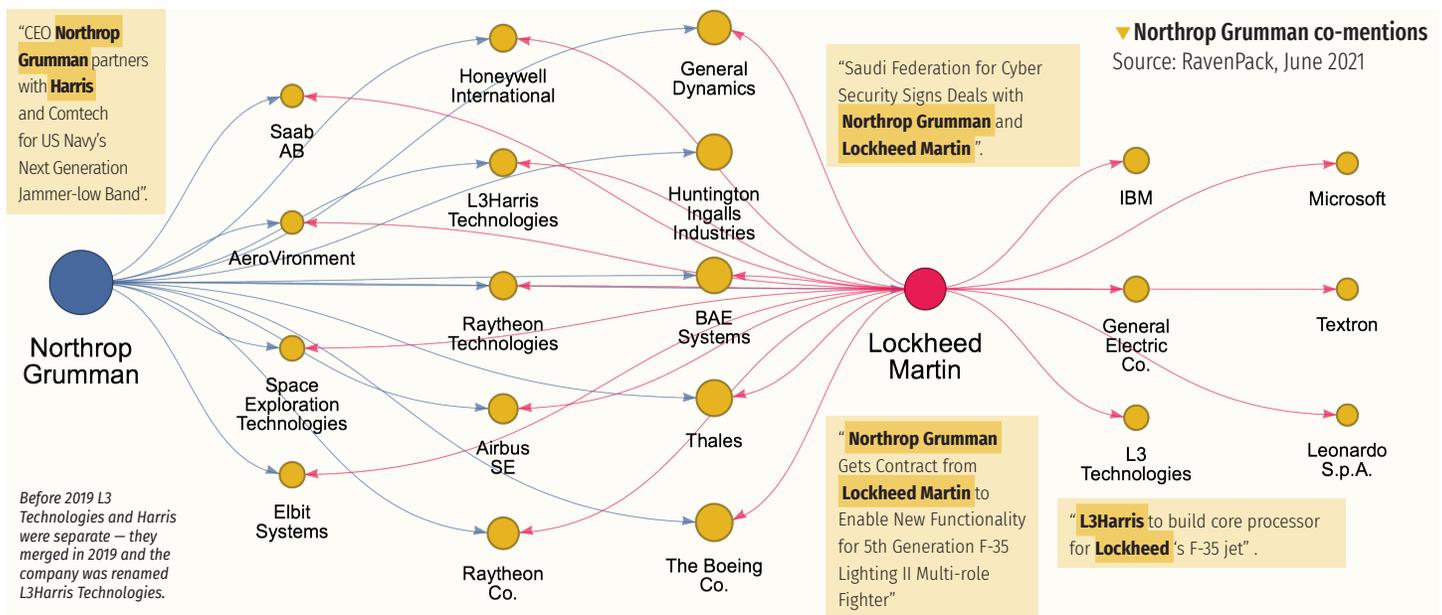
One approach is to start with a verified set of companies belonging to an exclusion list, then expand it with firms that appear in the same articles with high relevance. Tracking company co-mention volumes allows analysts to maintain such lists through fundamental links researched and verified by journalists.

To illustrate the concept, take Northrop Grumman — one of the world’s largest weapons manufacturers and military technology providers. The news co-mention analysis uncovers a network of predictable candidates, such as L3Harris Technologies, an American technology company specializing in surveillance solutions, microwave weaponry, and electronic warfare. Using

the network, we can quickly unearth the context behind the connection: Northrop partnered with L3Harris and Comtech to support the U.S. Navy’s next-generation jammer-low band program.

Unsurprisingly, Northrop is also linked with Lockheed Martin, the American aerospace, defense and advanced technologies company. Further context reveals that both companies have signed deals with the Saudi Federation for Cyber Security and separately unveiled a partnership to enable new functionality in the 5th generation F-35 fighter jet. By expanding the network, we can see that Lockheed further engages with players outside of the defense industry.

ESG investors can leverage these co-mention networks to screen for companies outside of aerospace and defense but tacitly linked to the industry. For example, a chip manufacturer engaging with the major players could raise a red flag from an ESG perspective.



# 9

## ANTICIPATE MARKET REACTIONS TO ESG ADVERSE MEDIA

Systematically monitoring for controversies across a portfolio empowers investment managers to react to changes in sustainability risks in a timely fashion. Analyzing the historical impact of similar events can also help managers decide how to manage the risk associated with evolving situations.

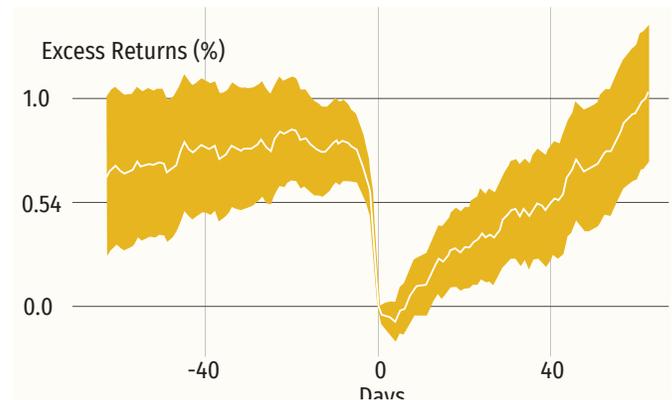
RavenPack analyzed the average performance of fraud and scandal events across the top 3,000 companies in the U.S. based on market capitalizations. RavenPack's research shows that stocks involved in such events took a significant intraday hit and continued trading lower for several days. However, they also observed an apparent overreaction in selling, leading to a price reversal over subsequent months.

The tendency of stocks to react uniformly creates a repeating market pattern that informed investors could use to respond to negative ESG news strategically. By adjusting their investment strategy, they can capitalize on these moves. For example, in the case mentioned above, it might be prudent to delay selling a stock that had just experienced an ESG controversy, as the price is likely to drift higher to more attractive levels.

Researchers at Monash University in Australia found the same market overreaction pattern with a subsequent drift higher following most other types of reputationally damaging ESG news. Their research included events like litigations, labor and regulatory issues, crime, security, pollution, lapses in corporate responsibility, and adverse health-related events.

To illustrate this dynamic from a portfolio construction perspective, an investment strategy that trades on the initial sell-off, with a holding period of 3 to 5 days, delivers annualized returns of more than 20% with an information ratio of 1.6 (see chart on the right).

An alternative version of the strategy was also found to capture the reversal signal to capitalize on the upward drift, trading one week after the adverse ESG event to avoid the previously mentioned momentum effect. In all these examples, news sentiment analytics provided a highly relevant signal that can help investors capture short-term opportunities to bolster the performance of their ESG portfolio.



▲ Schematized market reaction to headlines containing “fraud” or “scandal.” Source: RavenPack, June 2021



▲ Cumulative excess returns of strategy taking advantage of the short-term sell-off following “controversy” events with negative sentiment. Source: RavenPack, June 2021

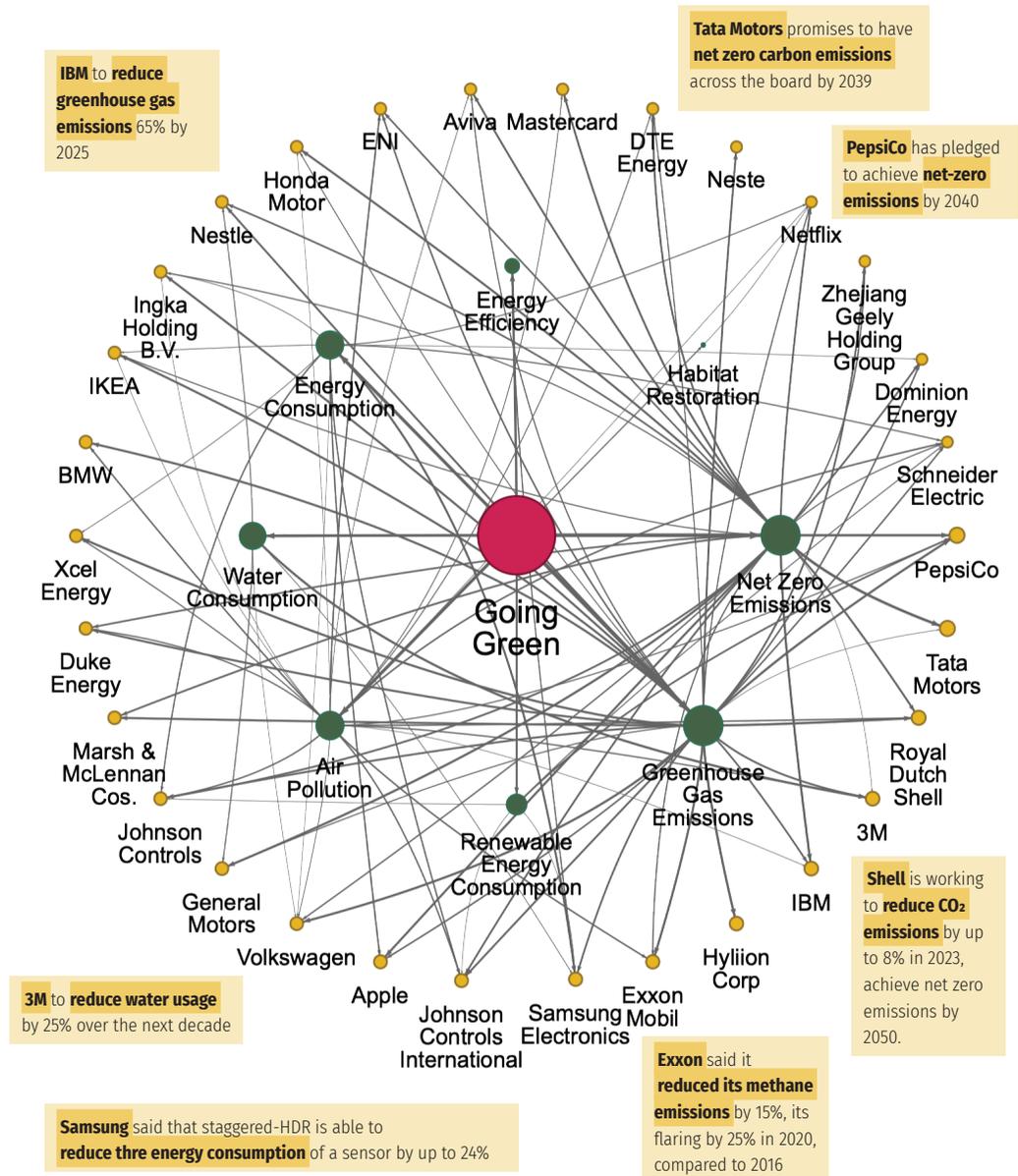
# 10

# ELEVATE FROM ESG KEYWORDS TO SUSTAINABLE THEMES

With the number of companies announcing sustainable initiatives on the rise, ESG investors look to identify those that are embracing the trend. Constructing news co-mention networks can offer a valuable starting point for identifying different companies linked to specific themes.

Defining a theme by identifying a set of key phrases that encapsulate a subject matter is relatively straightforward. However, this approach can be further refined by overlaying pre-defined ESG event taxonomies and event detection capabilities to improve precision. Alternatively, using semantic similarity search offers a more flexible means of topic detection by capturing broader context without the restrictions characteristic of keyword-type approaches.

The following graph shows a network of companies linked to the theme of going green. ESG investors can further refine the topic based on individual sub-themes or co-mention frequencies, and analyze time series data to identify new entrants as a basis for further research.





# RavenPack

Since its inception, RavenPack has focused on developing the technology to measure sentiment and identify actionable events systematically across news and social media content. Today, we cover more than 6,800 financial and non-financial event categories across 300,000+ entities (including companies, people, products, places, and organizations, among others).

Analytics are made available to our clients at a granular level, in real-time, across more than 25,000 news sources, with links to the underlying textual document for full transparency. In addition, the analytics can be further filtered and aggregated depending on end users' objectives and use cases.



Sales Team: [sales@ravenpack.com](mailto:sales@ravenpack.com)



[www.ravenpack.com](http://www.ravenpack.com)



Technical Support: [support@ravenpack.com](mailto:support@ravenpack.com)



[info@ravenpack.com](mailto:info@ravenpack.com)



Americas: **+1 646 277-7339**

EMEA: **+44 (0) 20 3290 7395**

APAC: **+65 3138 4809**