# **Analysis of Topic Trends in Biomaterial Research**

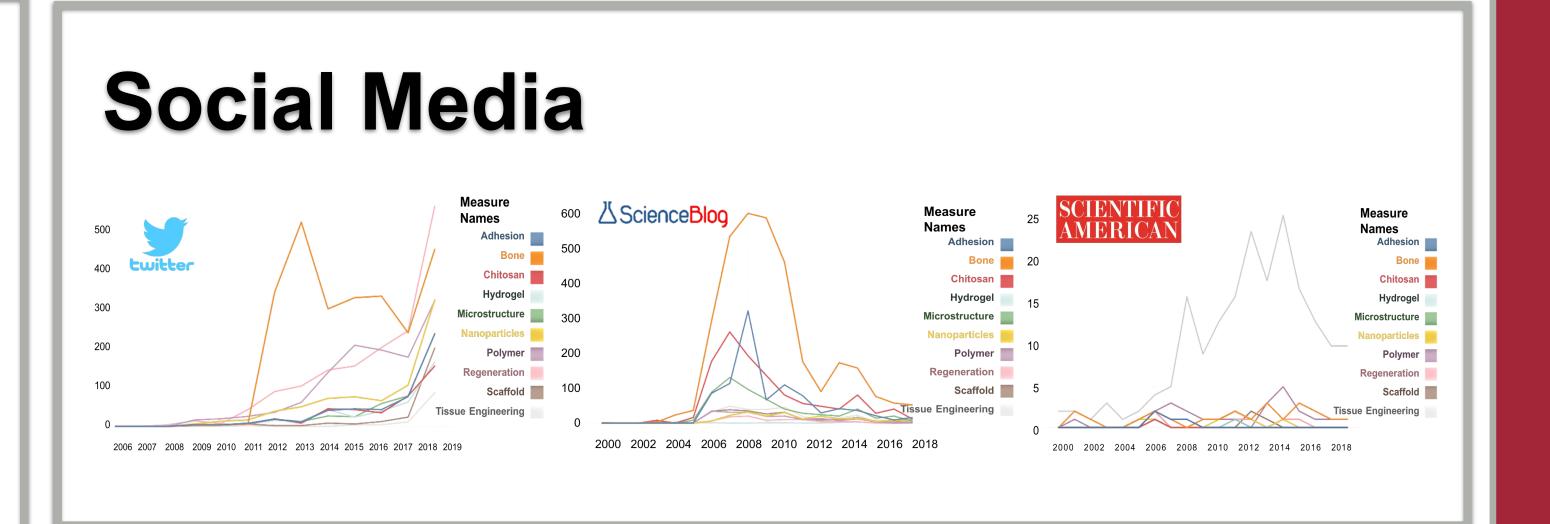
Instructor: Christopher Asakiewicz Jingsi Gong, Shivi Jain, Shuai Wang, Yuhan Hu



Business Intelligence & Analytics

# **Background & Motivation**

A great number of biomaterials have been strongly developed in many fields of academic research. For academic publishing companies like Weily, the prediction of topics trends can make contribution to the marketing strategy. Our objective is to study the topic trends in biomaterial research and investigate effective features which possibly indicate the emerging and shrinking of topics. Especially, the influence of social medias is discovered.

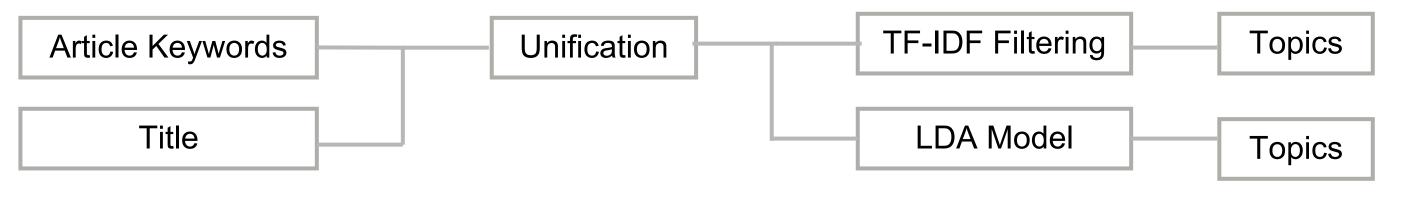


# **Topics Extraction Process**

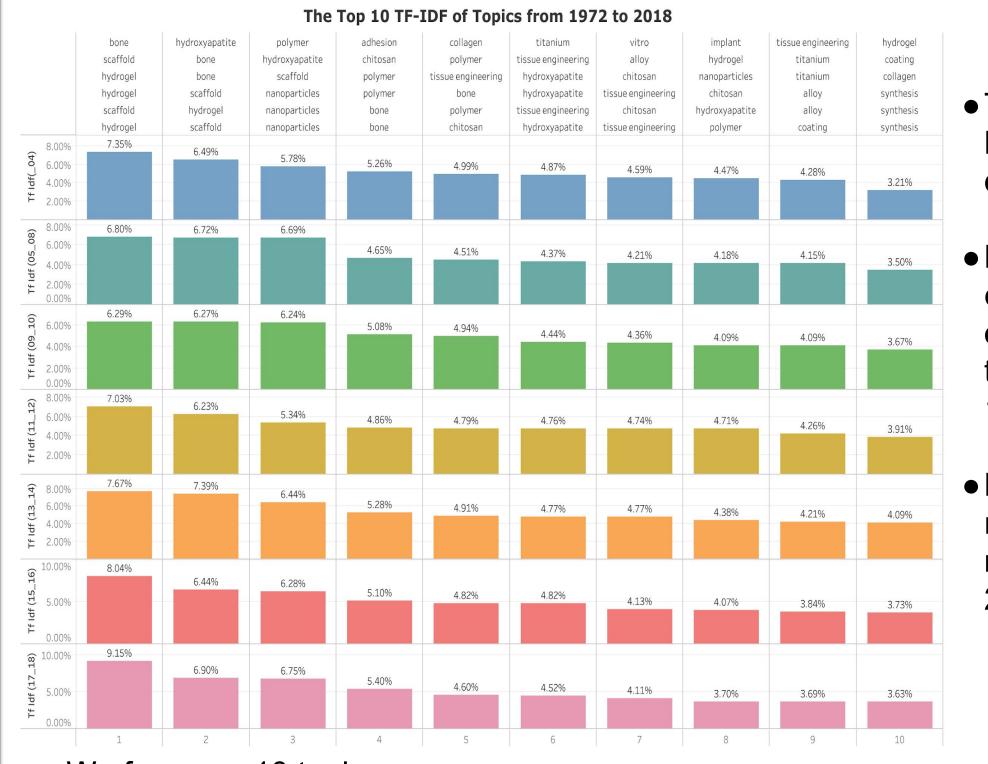
Data Source: Web of Science Database

Search Terms: "Biomaterials" or "Biomedical Materials"

Record Count: 43480 Timeline: 1972-2018



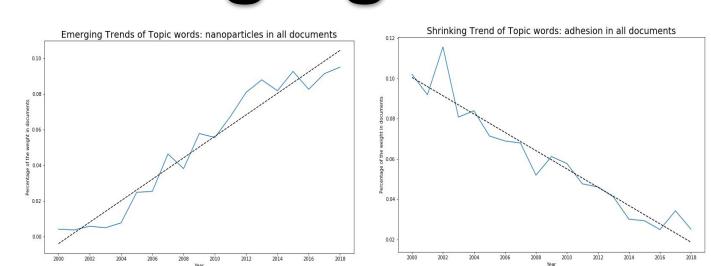
# **General Overview**



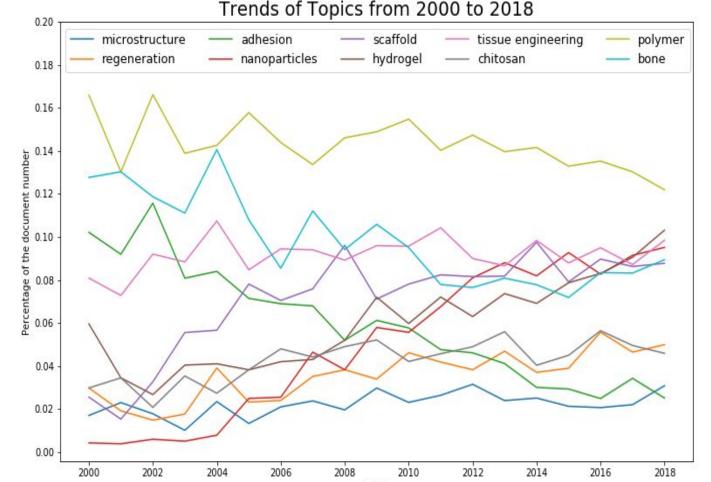
- Topics like bone,
  hydroxyapatite, polymer
  dominated the field before 2010
- Bone, hydrogel, polymer, chitosan and tissue engineering have stayed on the top 10 topics list for more than 10 years
- Hydrogel, scaffold and nanoparticles have emerged as rising hot topics since around 2009

We focus on 10 topics: Adhesion, Bone, Chitosan, Hydrogel, Microstructure, Nanoparticles, Polymer, Regeneration, Scaffold, Tissue Engineering

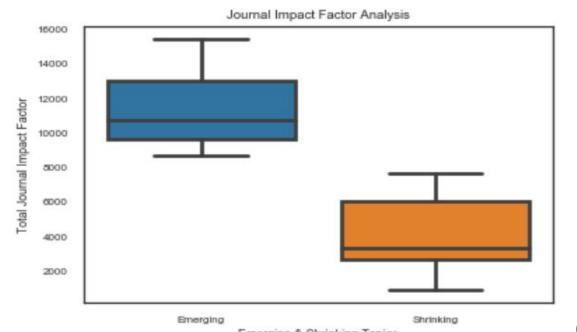
# **Emerging and Shrinking Analysis**



- The emerging trend is discerned with the positive coefficient of the linear fit, while the shrinking trend is correlated with the negative one.
- The emerging topics includes tissue engineering, scaffold, hydrogel, chitosan, regeneration, microstructure, nanoparticles, while the topics of polymer, bone and adhesion are shrinking.



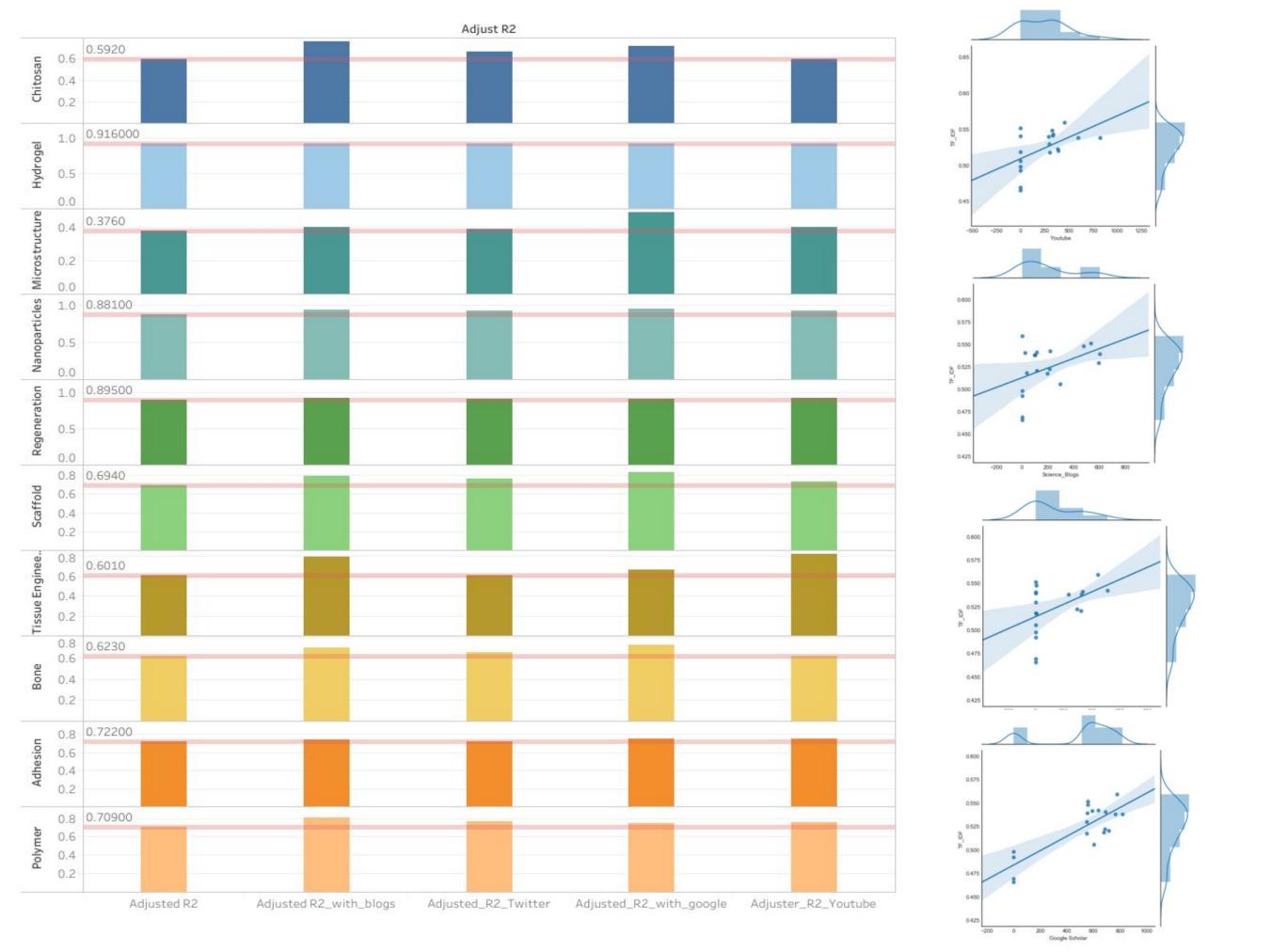
# Feature Analysis - Journal Impact Factor



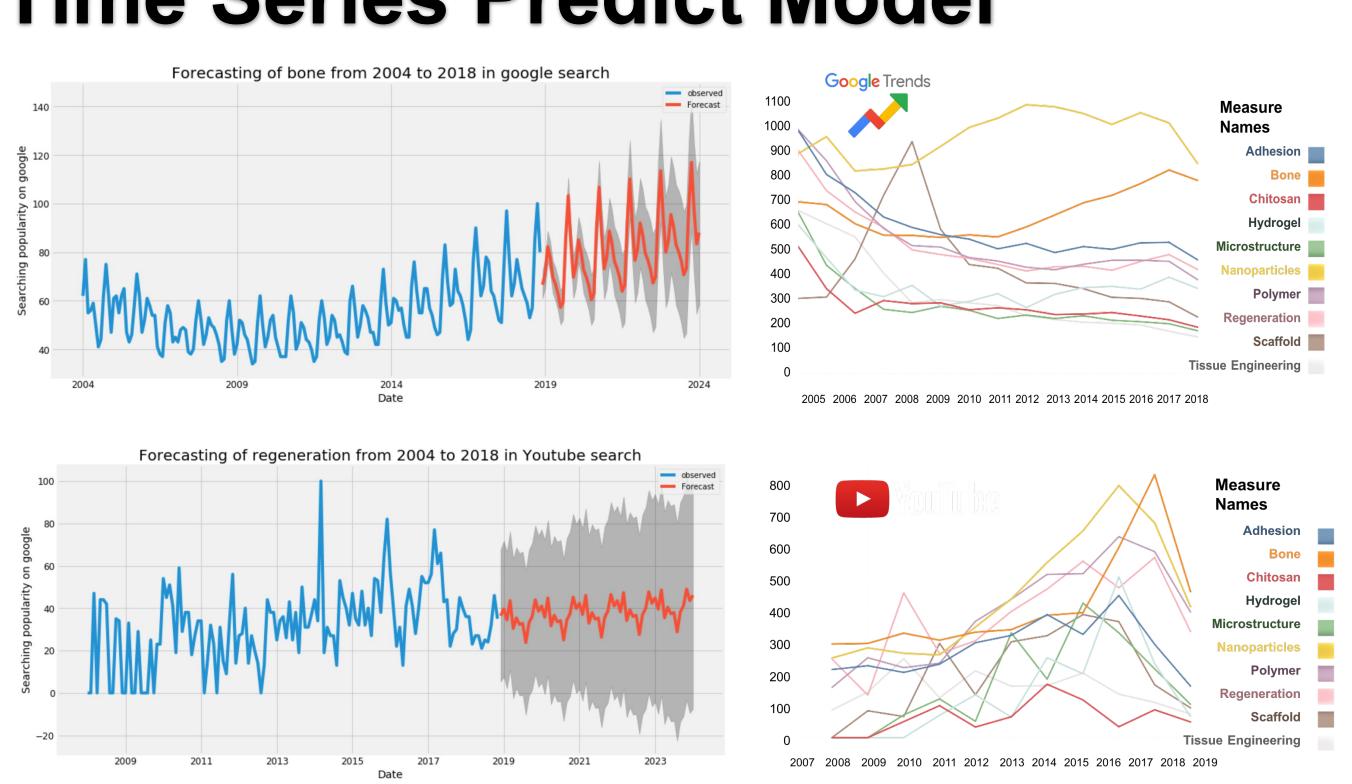
- Topics like Regeneration, Tissue, Biomaterials and Scaffolds, Tissue, Bone etc. have high Journal impact factor, which are also emerging topics as analyzed before.
- Topics like **Microscopy**, **Laser**, **DNA**, **beta** and strength etc. have very low Journal impact factor which are also shrinking topics.

(Here we using 24 topics that generated from LDA model)

# Feature Analysis – Social Media



# Time Series Predict Model



# **Conclusion and Future Work**

- We successively validated the previous teams' work on topic trend analysis in terms of two different method on generating topics.
- Both TF-IDF and LDA models affect the topic extraction. Also the update of database have a great influence on topic determination.
- The weight of topics in review portions has a positive correlation with the emerging trend of topics.
- The journal Impact factor, number of funding agencies, citations are strong indicators of emerging & shrinking trend.
- The social media is an effective factor to indicate the topic trends given by MLR model.
- The time series model approximately forecasted the emerging topics.
- In future we can append the geographical factor into our model. Also, updating univariate time series model to multivariate model could be an advanced method to improve the accuracy of prediction.

# Project Brief

## Group 1: Jingsi Gong, Shivi Jain, Shuai Wang, Yuhan Hu

## **Background and Motivation**

A great number of biomaterials have been strongly developed in many areas, such as: medicine, biology, chemistry, and materials science. For academic publishing companies like Weily, the prediction of topic trends can make contribution to the marketing strategy. Our objective is to study the topic trends in biomaterial research and investigate effective features which possibly indicate the emerging and shrinking of topics. Especially, the influence of social media was discovered and testified.

#### **Process and Conclusion**

## Topic Analysis and Validation

#### a. Overview of dataset

From Web of Science Database, we used keywords of biomaterials and biomedical materials to generate our datasets, which counts 43480 records in total from 1972 to 2018. Using VOSViewer to create density map, we had a general overview of hottest topics that consist of microstructure and regeneration.

## b. Topic Extraction

We utilized two methods to classify the topics. One is the TF\_IDF counting which is originated from the combination of keywords and titles. The other is LDA model that is able to create 24 topic groups including a series of topic words in each group. Both results are different from previous teams' conclusion respectively since the dataset has been changed over time. Also the algorithm of TF\_IDF and LDA made a difference on topic classification results.

### c. Feature Analysis

We explored five features including reviews, citation numbers, funding agency numbers, journal impact factors and publication types. All findings are consistent with that of previous teams.

## Social Media Analysis and Validation

### a. Scraped and classify the social media sources

The scraped social media can be divided into five different types, which contains google scholar, google trends, youtube search, twitters and blogs. And blogs compose of science blog, scientific american and Plos. For each type, we counted the number of articles or records in terms of topics. Due to the bias of different platforms in social media, there are significant differences among topic distributions.

#### b. Multi Linear Regression Model

Using TF\_IDF as the dependent variable in multi linear regression model, with independent variables of citation numbers, funding agency numbers and journal impact factors, we can assess the model by R^2 value to determine the relationship between TF\_IDF and other variables. Then added the variable of the social media, we generate R^2 value again to compare it with previous value. With increased R^2 value, we ascertained the effective impact of the social media on TF\_IDF that is articulated with the topic trends.

#### c. Time Series Model

Given testified impact of social media in indicating topic trends, we employed time series model to predict the future trend of topics based on google trends and youtube search. The results shows that the topics of **adhesion**, **polymer**, **bone** and **regeneration** are emerging in forecasting.

#### **Future Work**

In further improvement of model accuracy, we could also investigate and append the geographical factor into modeling. What's more, we can try multi time series model instead of using univariate time series model for the same purpose.