

Analysis of Beer Reviews

BeerAdvocate's Reviews



REALIZADO POR: Gabriel Simões nº59662 Jorge Caldeira nº58214 José Maria Diogo nº55937

Dados e Tendências do Mercado

O nosso foco foi compreender a estrutura subjacente dos dados, modelando relações entre reviews através de grafos.

Padrões de similaridade entre produtos e padrões de comportamento entre reviewers.





BeerAdvocate

O dataset original contém cerca de 1.5 milhões de reviews

- aroma
- appearance
- palate
- taste
- overall

Criação do Subset



01

Limpeza de Dados

Remoção de entradas sem teor alcoólico e substituição de perfis de revisores ausentes por "anônimo".



02

Seleção da Cervejaria

Foco na Boston Beer Company (Samuel Adams), com 38.812 avaliações de 7.251 revisores únicos.

03

Filtro Tópico

Restrição aos 10 estilos de cerveja mais avaliados, resultando em 21.012 avaliações.

Construção de Redes

1

Grafo de Similaridade de Revisão (S2)

Conecta avaliações com o mesmo estilo de cerveja e classificações idênticas em cinco aspetos (geral, sabor, aroma, aparência, paladar).

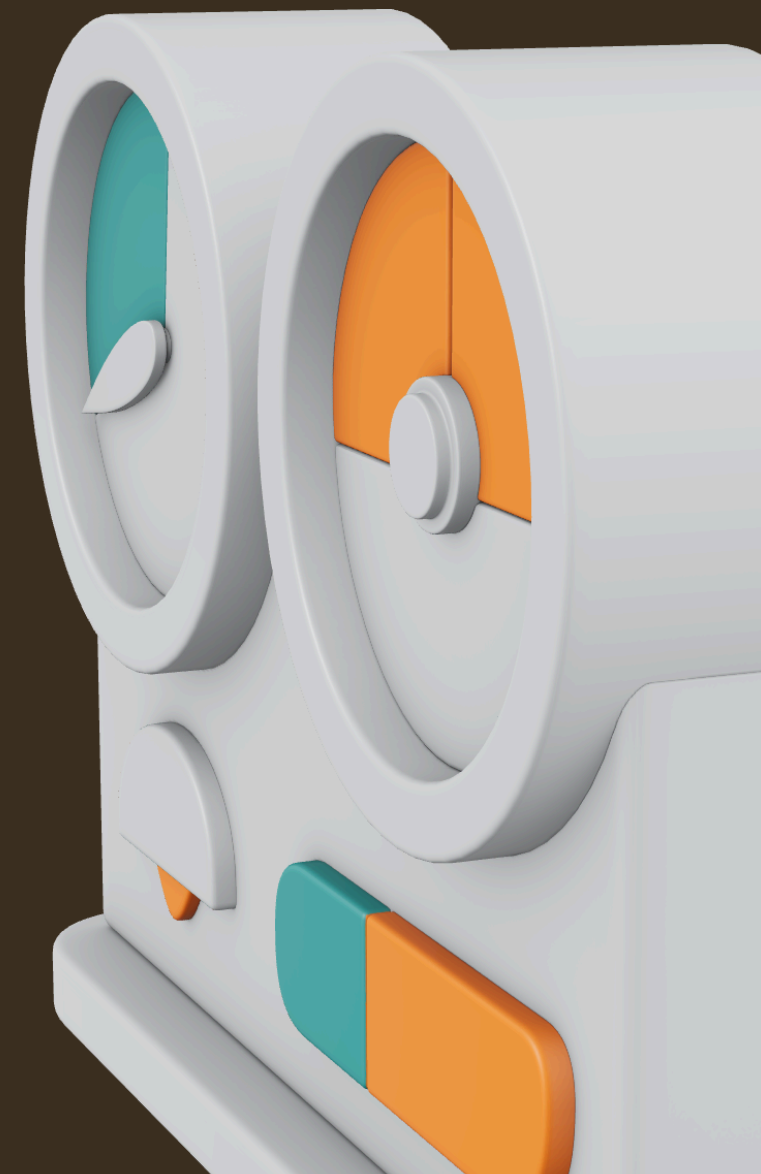
2

Grafo Baseado em Revisor (S3)

Conecta avaliações submetidas pelo mesmo revisor, modelando o envolvimento partilhado do utilizador.

Métodos de Análise

Métricas Globais Métricas de Centralidade Comunidades



Análise Comparativa

Ao analisar a estrutura global das duas redes, o padrão mais evidente é a extrema fragmentação.



Coeficiente de Agrupamento

Em ambos os grafos, o coeficiente médio de agrupamento é exatamente 1.



Estrutura de Cliques

Cada componente conectada é uma clique, implicando que todos os nós estão diretamente conectados entre si.



Conectividade Completa

A igualdade exata dos parâmetros força a conectividade completa entre todas as avaliações que partilham o mesmo perfil ou utilizador.

Grau e Centralidade

Distribuição de Grau

Grafo de similaridade: grau médio de 12, com alguns nós acima de 130.

Grafo de utilizador: grau médio de 7, com alguns nós acima de 25.

Nós de alto grau representam perfis de classificação extremamente frequentes.

Centralidade de Autovetor

A maioria dos nós tem valores próximos de zero, com alguns nós de alto valor próximos de 1.

A importância estrutural concentra-se nas maiores componentes.

Componentes maiores contribuem mais para os autovetores principais.

Densidade e Comparação

1

Modularidade

Valores muito altos, indicando que cada componente é uma comunidade densa e isolada.

2

Comprimento Médio do Caminho

Igual a 1, devido à estrutura de cliques completas onde qualquer par de nós conectados está diretamente ligado.

3

Densidade

Ambos os grafos são extremamente esparsos, refletindo alta fragmentação.

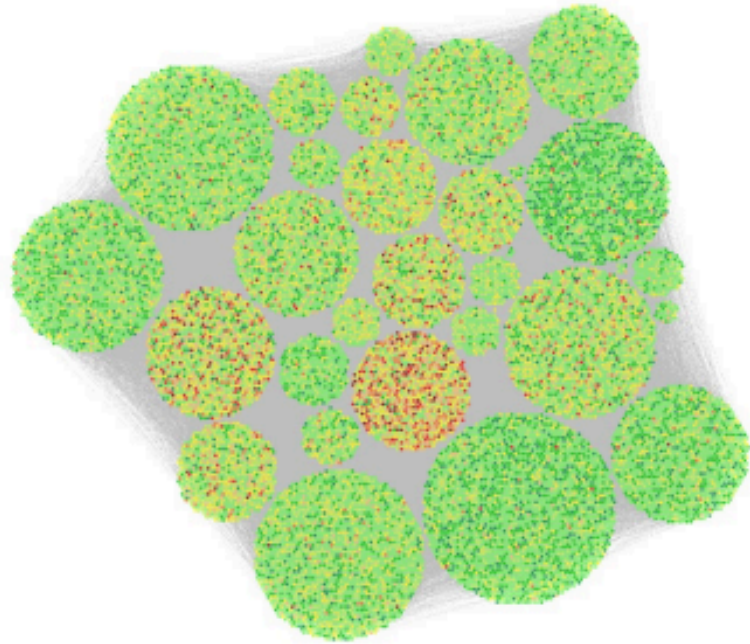


Figure 1: Reviews connected by multiple review parameters in common

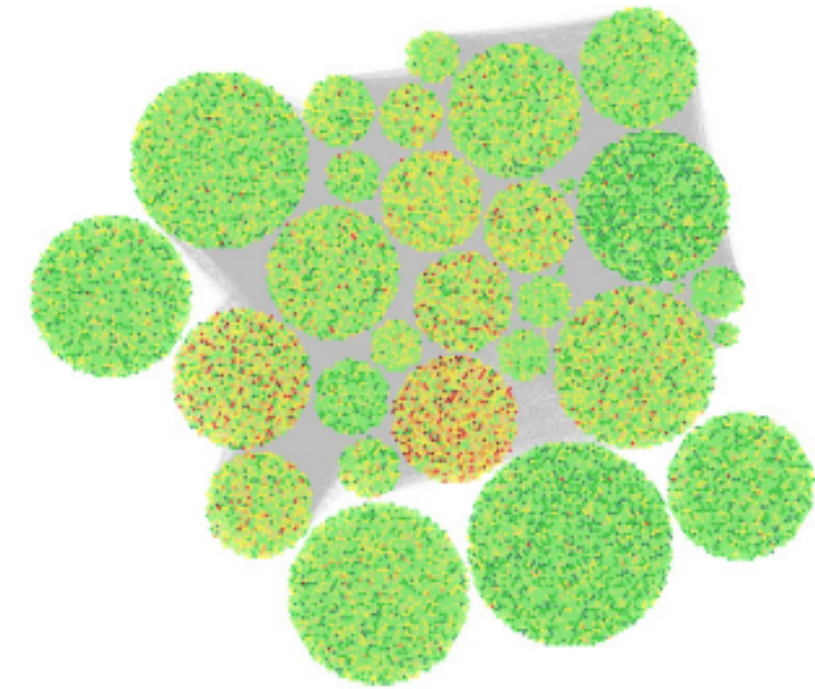


Figure 2: Reviews connected by reviewers in common

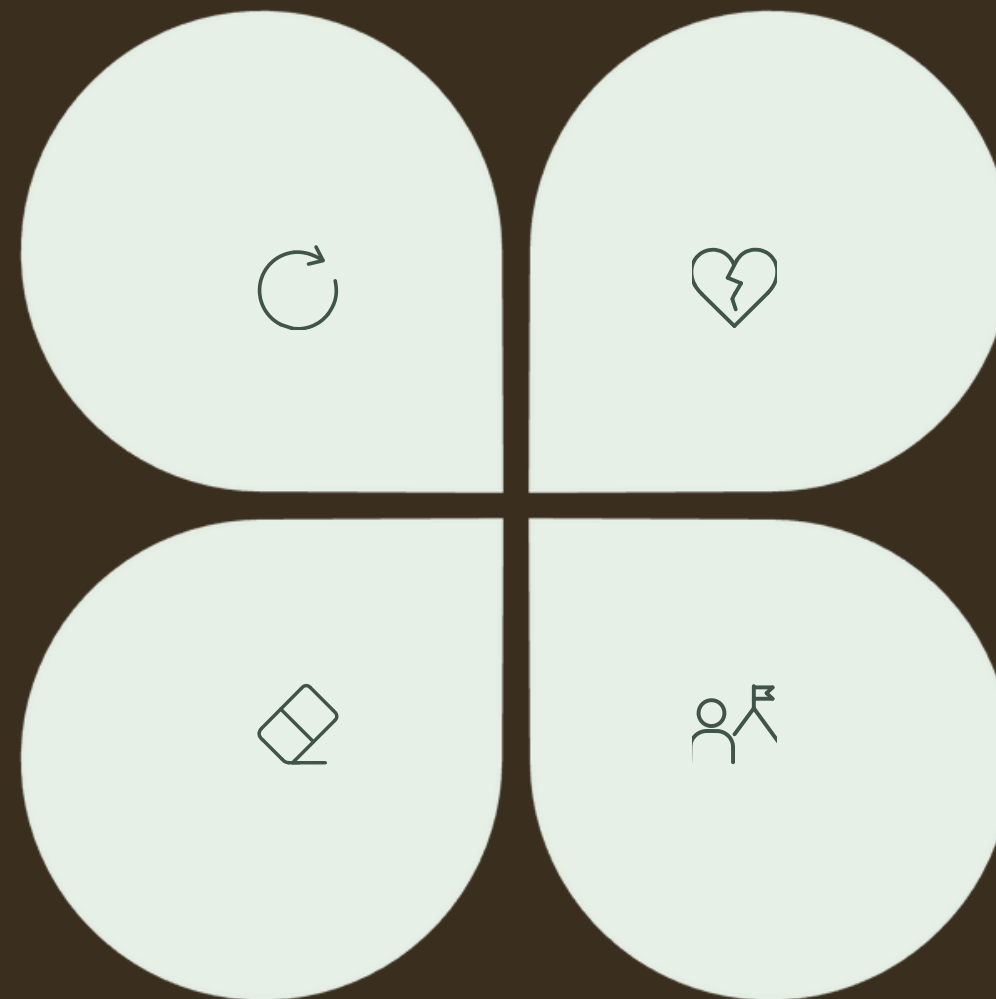
Discussão e Limitações

Redundância

Componentes como cliques completas implicam redundância significativa de arestas.

Nuances

Tratar componentes como cliques completas remove nuances estruturais.



Fragmentação

Densidades globais muito baixas refletem extrema fragmentação.

Desafios

Analisar grandes conjuntos de dados através da teoria de redes apresenta desafios computacionais.



Conclusão e Referências futuras