



A ESTRUTURA DA WEB

Redes Sociais e Econômicas

Prof. André Vignatti

A ESTRUTURA DA WEB

até agora: redes onde unidades eram **peessoas** ou **entidades sociais**, como empresas e organizações

agora (Cap 13, 14 e 15), veremos as **redes de informação**, onde as unidades básicas são **pedaços de informação**

a **WWW** é o melhor exemplo de rede de informação

A ESTRUTURA DA WEB

mesmo com diferenças entre redes de informação e redes sociais e econômicas, **muitas ideias já discutidas** podem ser aplicadas

usaremos **Teoria dos Grafos** para discutir a estrutura da Web e mecanismos de busca na Web

usaremos **leilões, mercados e teoria dos jogos** para discutir sobre publicidade em mecanismos de busca

WORLD WIDE WEB

I teach a class
on Networks.

Networks
Course:
We have a
class blog

Networks
Class Blog:
This blog post
is about
Microsoft

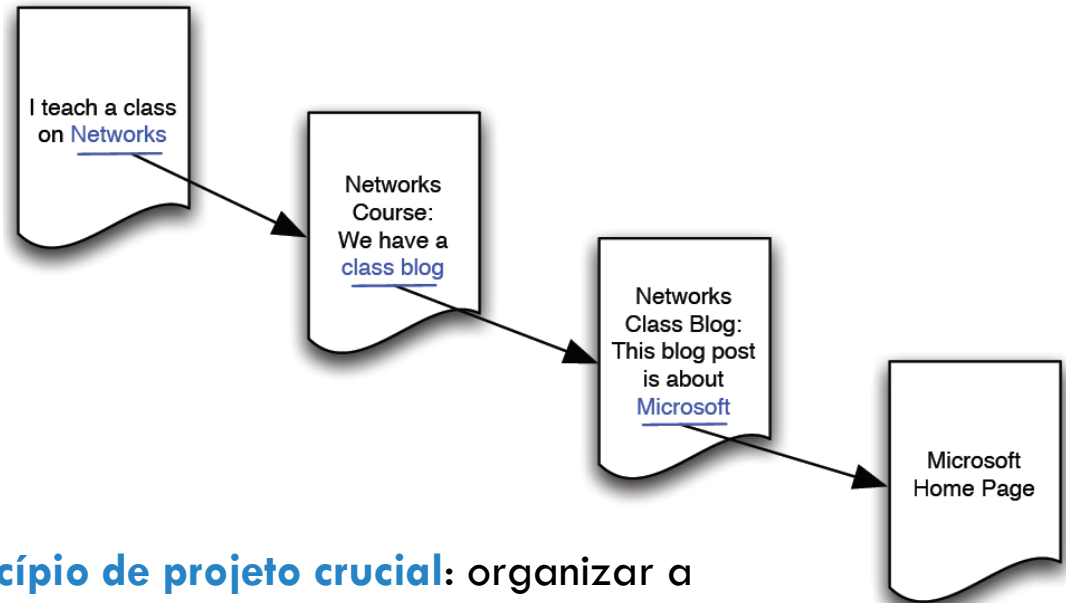
Microsoft
Home Page

a **WWW** é um **aplicativo** desenvolvido para “*permitir as pessoas compartilhar informação sobre a Internet*” criada por **Tim Bernes-Lee** durante **1989** a **1991**

duas principais características:

1. **páginas Web**: um meio de **tornar documentos disponíveis** a qualquer um na Internet
2. **browser**: um meio de outros **acessarem as páginas Web**, conectando-se computadores públicos na Internet e recuperando as páginas armazenadas neles

HIPERTEXTO



além disso, há um **princípio de projeto crucial**: organizar a informação usando uma **estrutura de rede**

ao criar uma Página Web, é possível **anotar uma parte do documento** com uma **ligação virtual** a outra Página Web assim, as Páginas Web viram uma “**web**” de Páginas Web

HIPERTEXTO

interpretação como grafos direcionados:

- os **nodos** são as páginas
- se a página **A** tem uma ligação para a página **B**, então há uma **aresta direcionada** de **A** para **B**

HIPERTEXTO

uma ideia genial: decidir organizar a Web como uma rede é uma **ideia inspirada e não-óbvia!**

- existem muitas maneiras de organizar a informação:
 - **sistemas de classificação** (livros numa biblioteca)
 - **várias pastas** (arquivos no computador)
 - **alfabeticamente** (como lista telefônica)

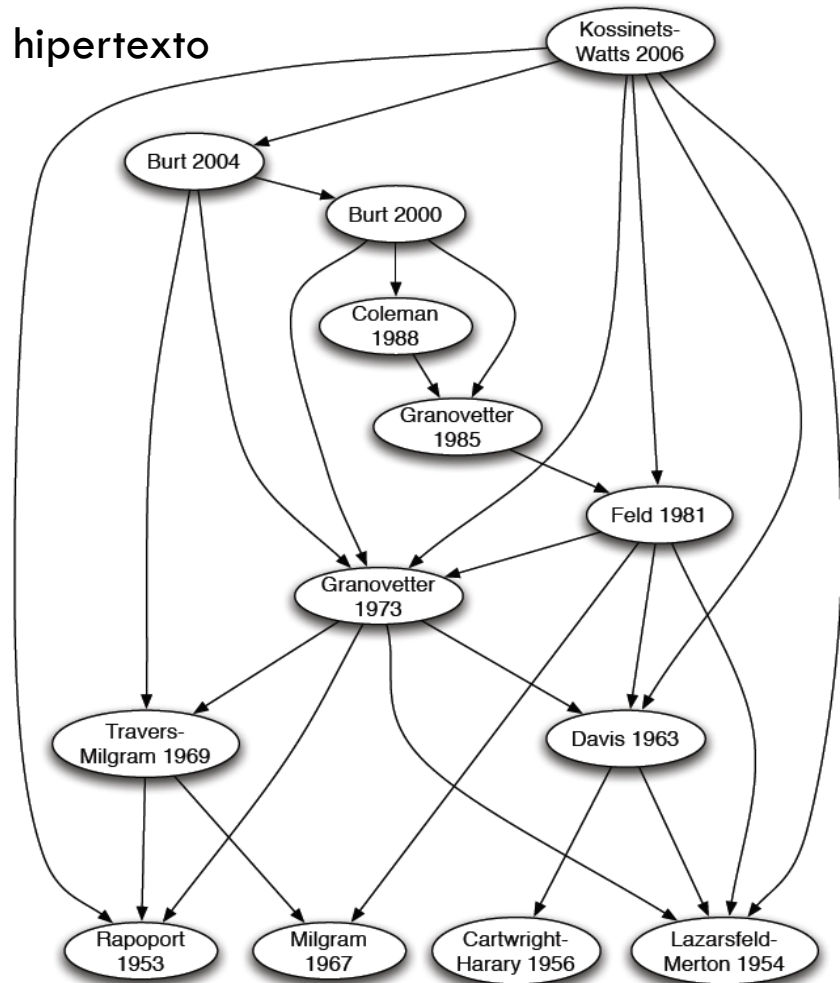
a ideia de hipertexto remonta a **estudos da metade do século 20**

REDES DE INFORMAÇÃO E HIPERTEXTO

redes de citações acadêmicas: precursor do hipertexto

diferença entre rede de citações e WWW:

- nas citações, as arestas são governadas pelo **tempo cronológico**
- na WWW, as páginas antigas podem ser **modificadas**

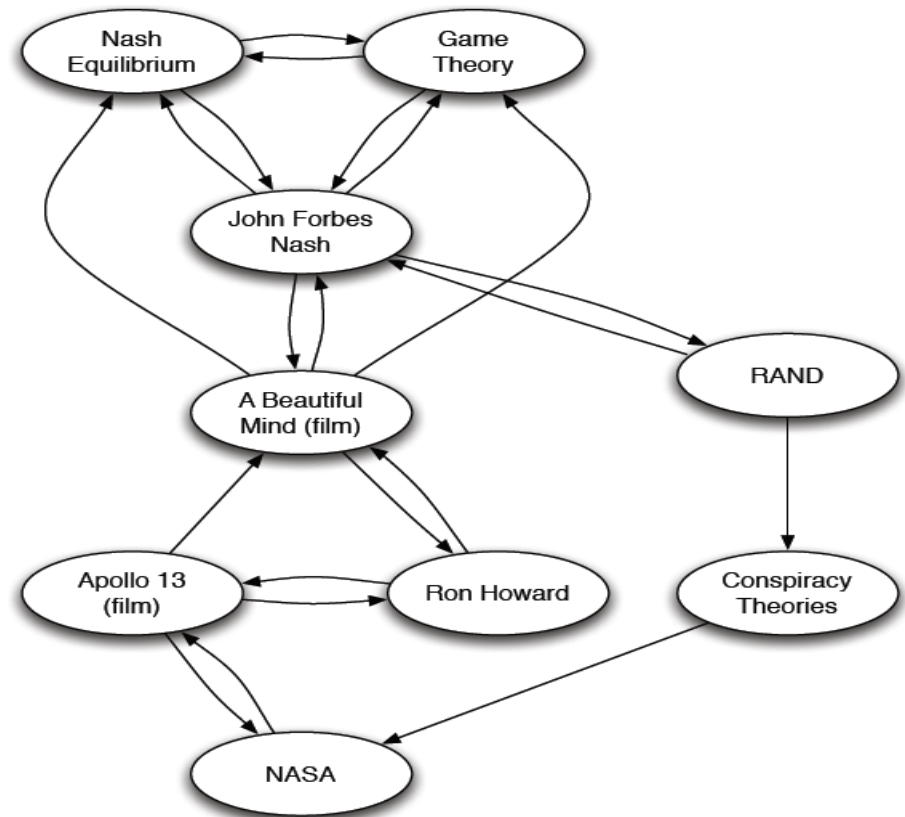


REDES DE INFORMAÇÃO E HIPERTEXTO

outro precursor: [enciclopédias](#)

na figura, um exemplo de ligações entre **tópicos de Teoria dos Jogos** na **Wikipedia**

note que a Wikipedia pode ter [mudado com o tempo](#)



REDES DE INFORMAÇÃO E HIPERTEXTO

primeiras ideias: Vannevar Bush - As We May Think
artigo de 1945 previa como a **computação e comunicação** iriam
revolucionar o acesso, troca e armazenamento de informação

ideia: os métodos de armazenar informações em livros, bibliotecas, memória do computador eram **lineares - item ordenados numa ordem sequencial**

REDES DE INFORMAÇÃO E HIPERTEXTO

mas Bush notou que nossa **experiência consciente de pensar** exibe uma propriedade chamada **“memória associativa”**

- você pensa uma coisa, aquilo faz você lembrar de outra, e assim por diante

ele propôs um **modelo de organizar informação** por links como na memória associativa

era algo parecido com a Web hoje: Tim Bernes-Lee disse que Vannevar Bush inspirou ele

EVOLUÇÃO DA WEB

no início: links que levavam a outras páginas Web

- com o passar do tempo, outros tipos de links surgiram: *“Adicione no carrinho”*, *“Submeter minha Pesquisa”*, *“Upload de Imagem”*, etc...
- tais links **NÃO** tem o objetivo de transportar a uma nova página Web

então os links são divididos em **duas categorias:** de navegação e de transação

- às vezes a divisão não é tão clara: *alguns links servem a ambas funções*

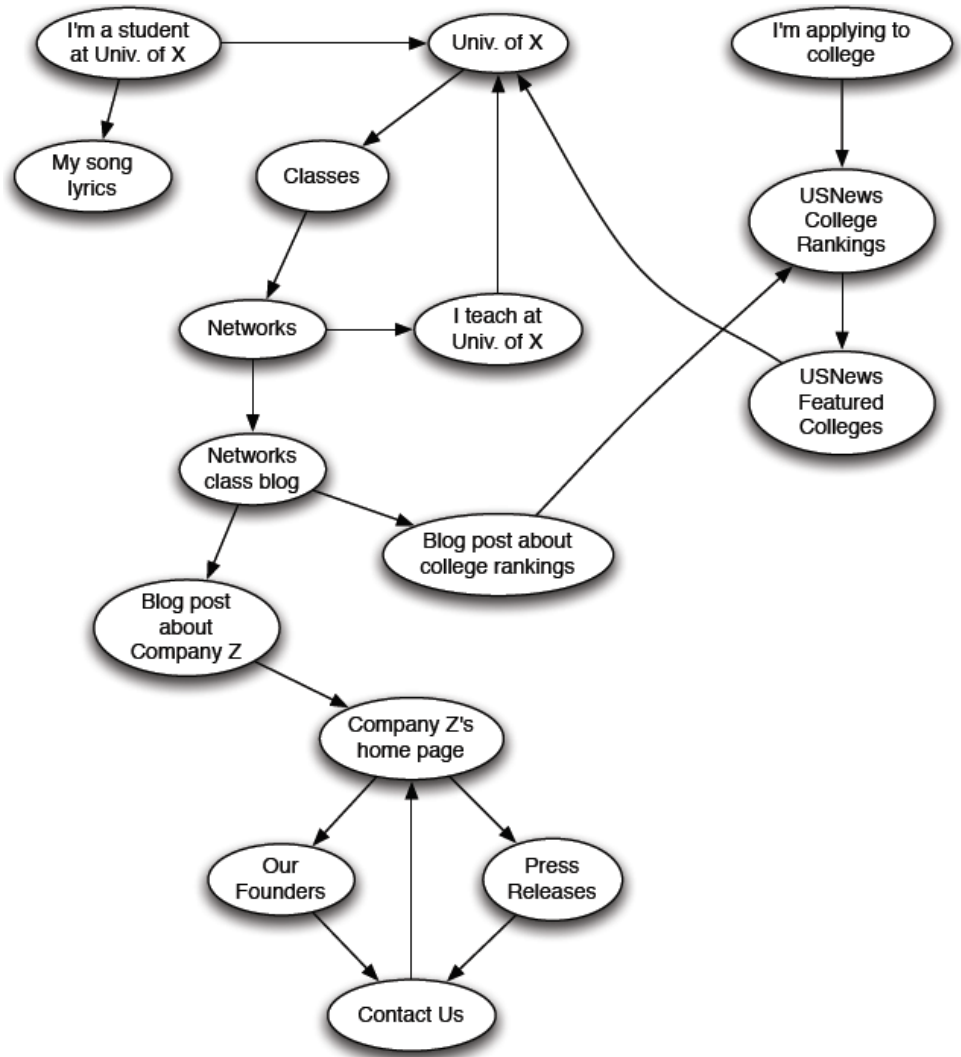
no que segue, focamos somente em links de navegação

A WEB COMO UM GRAFO DIRECIONADO

olhar **somente links de navegação**
ainda **fornece o núcleo estrutural** da
Web

o fato da **natureza direcionada**
torna a Web *diferente dos grafos*
visto até agora

- em redes sociais, **não há**
razão de usar direção



A WEB COMO UM GRAFO DIRECIONADO

relembrando (grafos não direcionados):

a **conectividade** foi definida **em termos de caminhos**:

- um grafo é **conexo** se todo par de nós está ligado por um caminho

definição para **grafos direcionados (www)**:

caminho: um caminho de um nó **A** até nó **B** é uma **seqüência de nós** que começa com **A** e termina com **B**, onde cada par consecutivo de nós é ligado por **uma aresta que aponta na direção de avanço**

A WEB COMO UM GRAFO DIRECIONADO

grafo fortemente conexo: se há um caminho (**direcionado**) de todo nó para todo nó

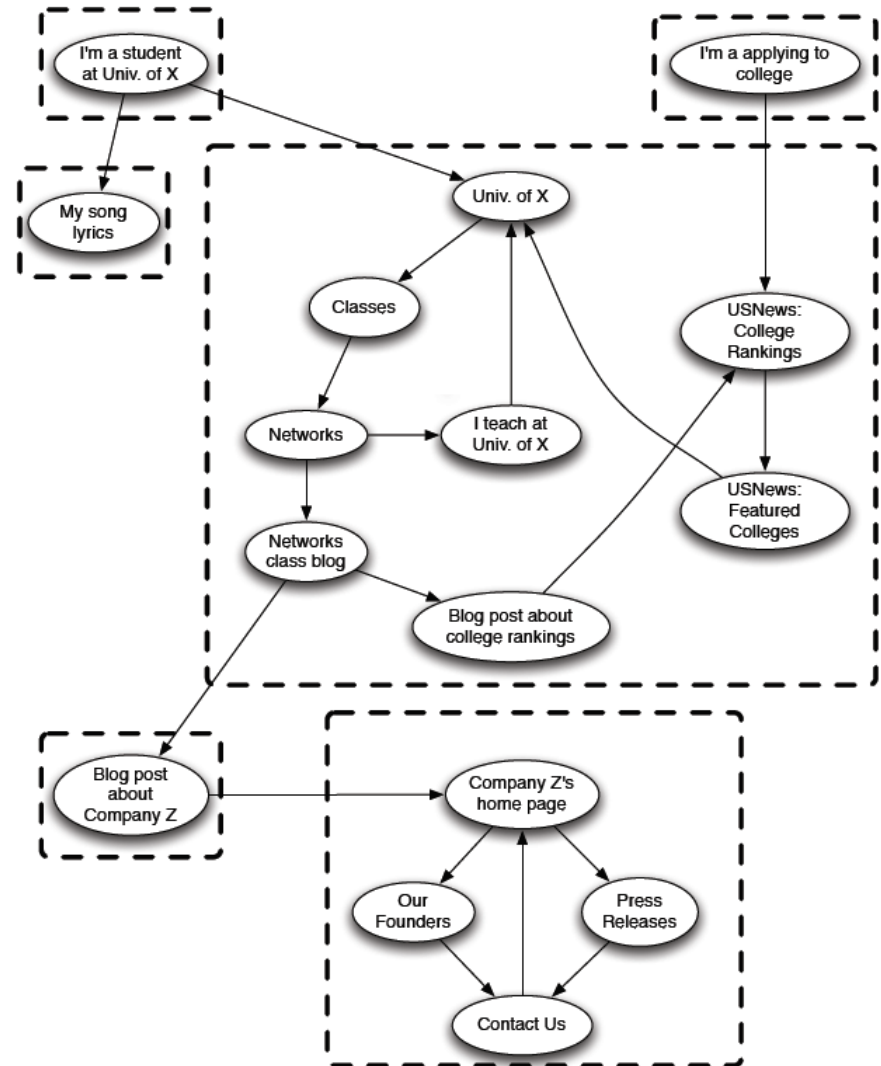
nem sempre um grafo é fortemente conexo, às vezes só **algumas partes do grafo** são assim

isso nos leva a definição de “**componentes fortemente conexas**”

A WEB COMO UM GRAFO DIRECIONADO

um **componente fortemente conexo** em um **grafo direcionado** é um subconjunto dos nós tal que:

- 1) cada nó do subconjunto tem um **caminho para todo outro nó**; e
- 2) o subconjunto **não é parte de algum conjunto maior** com a propriedade (1)



A ESTRUTURA DA WEB

como seria um “**mapa**” da web?

- obviamente, não seria como um mapa do mundo físico, dado o tamanho e complexidade de rede
- queremos algo mais “*abstrato*”, que mostra de **maneira estilizada** como seria a **estrutura da Web**

ideia: dividir o “mapa” em algumas **partes grandes**, e mostrar como **essas partes se encaixam**

COMPONENTE FORTEMENTE CONEXO GIGANTE

descobertas inicial: a Web tem um **componente fortemente conexo gigante**

ideia: páginas de busca ou “páginas de início” tem links para sites grandes e importantes, que levam a grandes instituições, empresas, universidades e órgãos do governo

- cada página dessa, tem os links para as **subpáginas menores dentro do mesmo domínio**
- cada página menor geralmente tem um **link para a página grande do domínio**, ou para **motores de busca**

assim, tais páginas **se alcançam mutuamente**, e portanto pertencem ao **mesmo componente** fortemente conexo

COMPONENTE FORTEMENTE CONEXO GIGANTE

dado que esse componente contém pelo menos páginas das maiores organizações comerciais, governamentais e ONGs, é fácil perceber que esse componente é grande

existem dois componentes conexos gigantes?

RESPOSTA: não. Se existisse dois componentes fortemente conexos gigantes X e Y , bastaria uma aresta de X para Y e outra de Y para X para que o componente se tornasse um só

A ESTRUTURA DE “GRAVATA BORBOLETA”

como posicionar os outros componentes fortemente conexos em relação ao componente gigante?

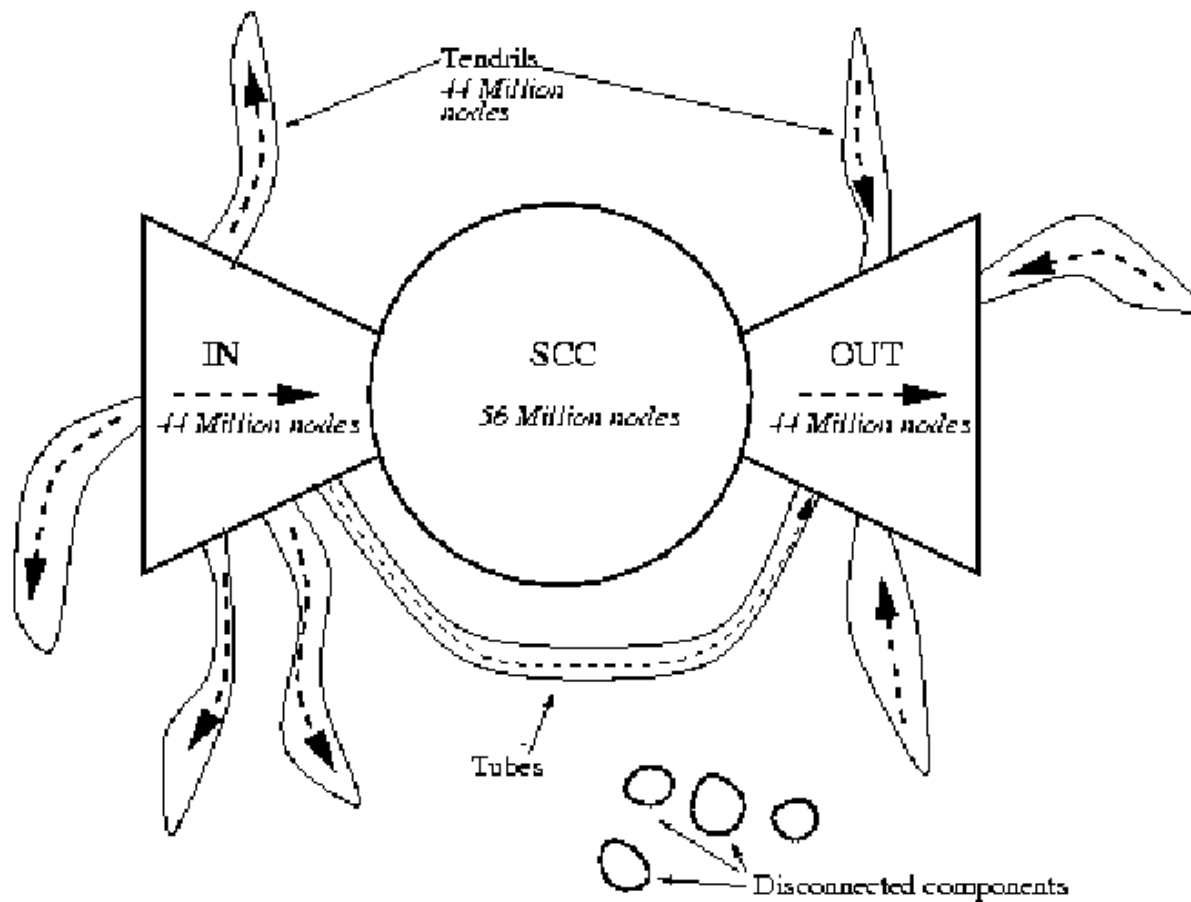
- **IN:** nós que **podem alcançar** o componente gigante
- **OUT:** nós que **são alcançados** a partir do componente gigante

A ESTRUTURA DE “GRAVATA BORBOLETA”

como posicionar os outros componentes fortemente conexos em relação ao componente gigante?

- **tendrils** (gavinhas):
 - I. Nós que são alcançados a partir de IN e não alcançam o componente gigante
 - II. nós que alcançam OUT e não podem ser alcançados a partir do componente gigante
- **tubos**: é um tendril que satisfaz (I) e (II)
- **componentes desconectados**: nós que não tem caminho para o componente gigante, mesmo ignorando a direção das arestas

A ESTRUTURA DE “GRAVATA BORBOLETA”



O SURGIMENTO DA WEB 2.0

na **2ª década da Web (2000-2009)** surgiram as seguintes **tendências**:

- I. ferramentas para **criar conteúdo Web** e manter **conteúdo compartilhado** (blogs, Wikipedia)
- II. mover dados pessoais de seus próprios computadores para os **serviços em nuvem** (e-mails, fotos, vídeos)
- III. conexões on-line **entre as pessoas**, não apenas entre documentos (redes sociais)

A Web 2.0 é, principalmente, **“uma atitude, não uma tecnologia”**

O SURGIMENTO DA WEB 2.0

alguns exemplos das tendências:

- **wikipedia:** pessoas abraçaram a ideia de **edição de artigos em conjunto** criando uma enciclopédia aberta na Web (*princípio (I)*)
- **gmail** e outros serviços de e-mail incentivou as pessoas a deixar as empresas como o Google **hospedar seus arquivos de e-mail** (*princípio (II)*)
- **mySpace** e **facebook** tiveram uma adoção generalizada, com um conjunto de características que enfatizou principalmente a **criação de redes sociais online** (*princípio (III)*)

O SURGIMENTO DA WEB 2.0

outros casos usam mais de um princípio:

- **flickr, picasa e youTube**: lugar centralizado para armazenar fotos e vídeos (*princípio (II)*), as pessoas podem comentar e marcar as fotos e vídeos (*princípio (I)*) e forma conexões sociais entre as pessoas que seguem o conteúdo (*princípio (III)*)
- o **Twitter** *extendeu o princípio (II)*: armazena na *nuvem* experiências, pensamentos e questões de uma forma que antes não existia

FENÔMENOS SOCIAIS DA WEB 2.0

software que **melhora** a medida que **mais pessoas usam** (cap 16, 17, 19):

- Tradutor do Google
- Android x iOS

a **sabedoria das multidões** (cap 16 e 22):

- wikipedia sabe tudo
- melhoria da qualidade da notícias pelo Digg
- fotos de notícias frescas antes no Facebook, Picasa e Flickr do que nos meios de comunicação tradicionais

a **“cauda longa” – “the long tail”** (cap 18)

- por que o futuro dos negócios é **“vender menos de mais”**
- modelo de negócios da Amazon e Netflix

FENÔMENOS SOCIAIS DA WEB 2.0

estrutura de redes sociais grandes (cap 20):

- Facebook, MySpace, Orkut, etc... possibilitaram estudar estrutura de redes grandes

sistemas de reputação e confiança (cap 22):

- cap 5 vimos como bom (mau) comportamento afeta o **balanceamento estrutural**
- o papel ao fornecer informação essencial os **mercados online**

sistemas de recomendação (cap 18)

- guiar usuários a **conteúdos novos**
- papel ao distribuir a popularidade de itens e nos **mercados de nicho**