Monitoring Network Structure and Content Quality of Signal Processing Articles on Wikipedia

Tao C. Lee and Jayakrishnan Unnikrishnan

LCAV, EPFL
ICASSP, Vancouver, Canada
{tao.lee, jay.unnikrishnan}@epfl.ch

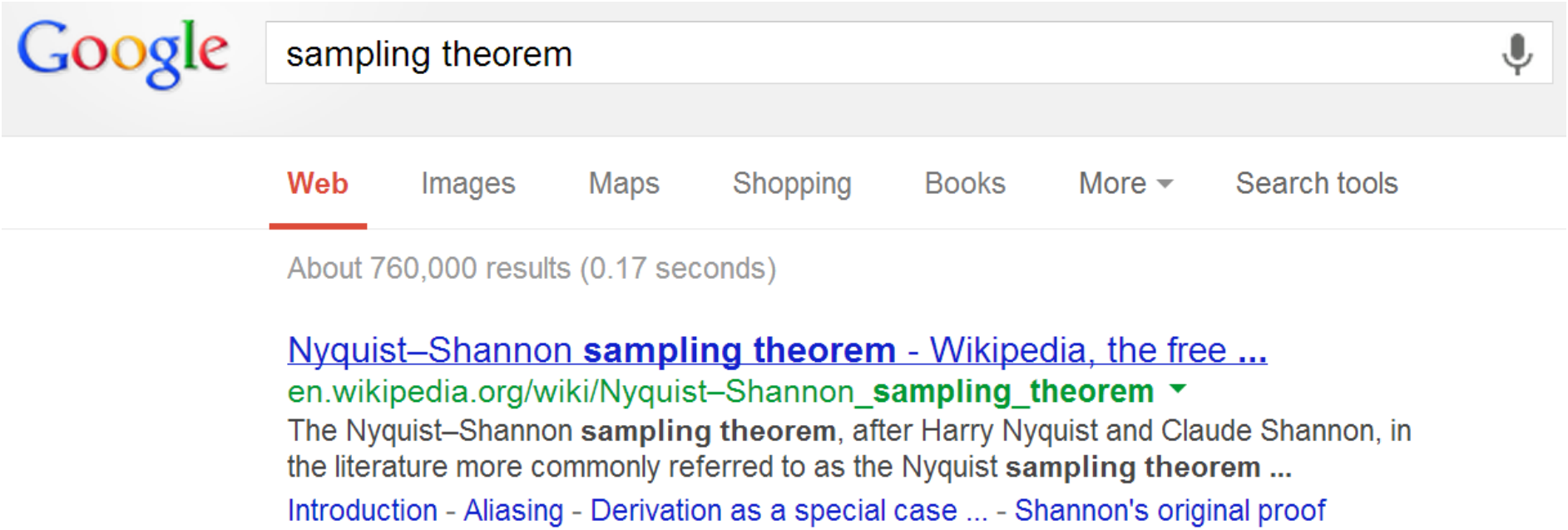
May 31, 2013

AUDIOVISUAL COMMUNICATIONS LCAV

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

A Google User's Impression

- Searching for **sampling theorem** on Google ...



A Google User's Impression

- An article with rich information ...

Nyquist–Shannon sampling theorem

From Wikipedia, the free encyclopedia

The **Nyquist–Shannon sampling theorem**, after Harry Nyquist and Claude Shannon,[1] in the literature more commonly referred to as the **Nyquist sampling theorem** or simply as **the sampling theorem**, is a fundamental result in the field of information theory, in particular telecommunications and signal processing. Sampling is the process of converting a signal (for example, a function of continuous time or space) into a numeric sequence (a function of discrete time or space). Shannon's version of the theorem states:[2]

> If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart.

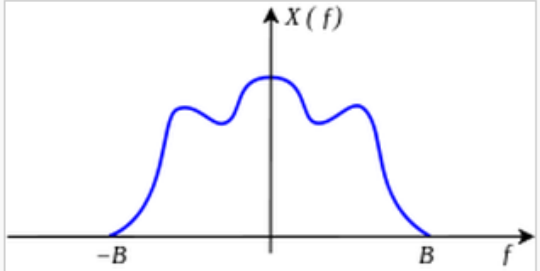


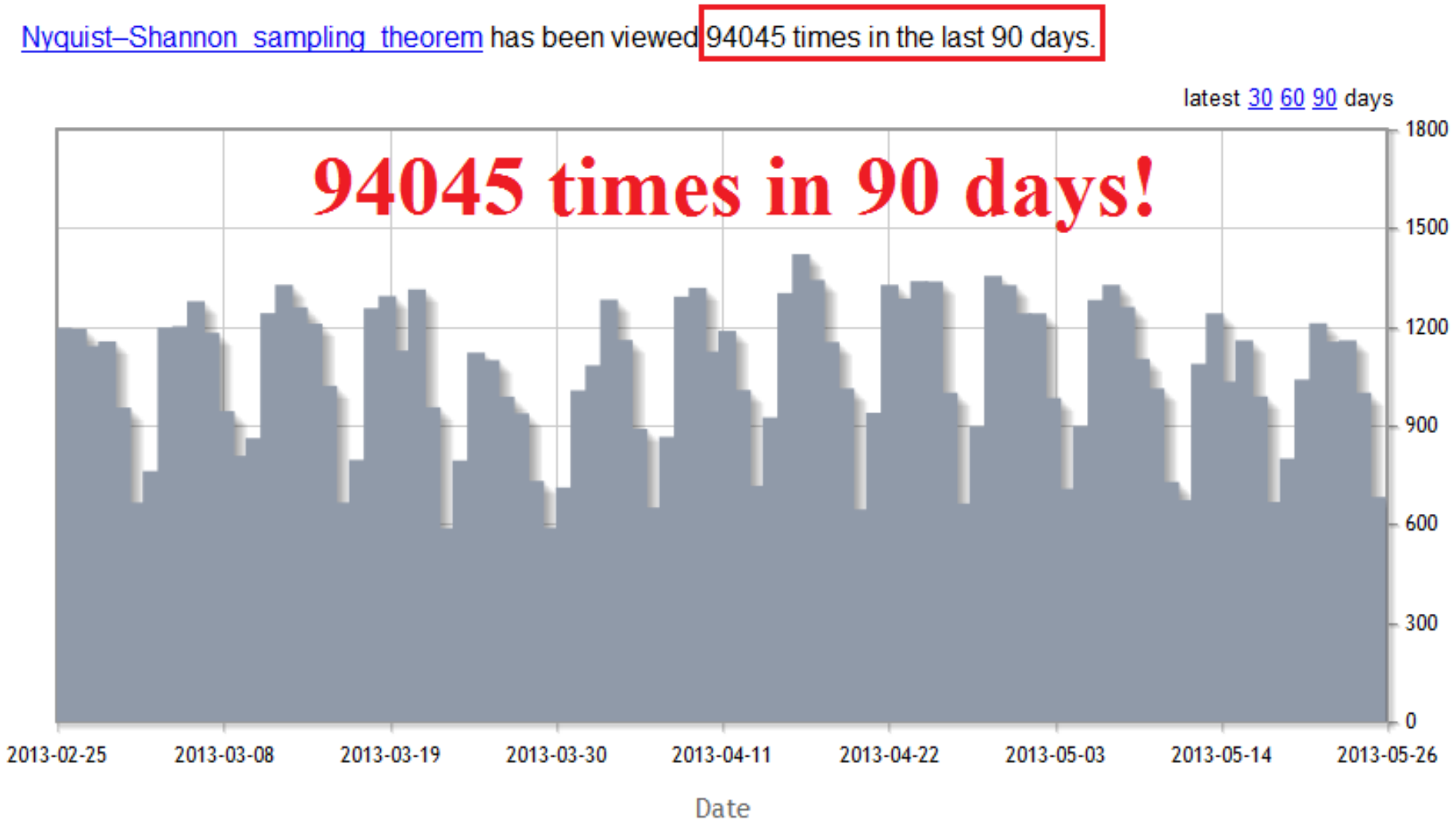
Fig. 1: Fourier transform of a bandlimited function (amplitude vs frequency)

In other words, a bandlimited function can be perfectly reconstructed from a countable sequence of samples if the bandlimit, B, is no greater than ½ the sampling rate (samples per second). The theorem also leads to a formula for reconstruction of the original function from its samples. When the bandlimit is too high (or there is no bandlimit), the reconstruction exhibits imperfections known as aliasing. The Poisson summation formula provides a graphic understanding of aliasing and an alternative derivation of the theorem, using the perspective of the function's Fourier transform.

In practice of course, infinite sequences, perfect sampling, and perfect interpolation are all replaced by approximations that deviate from the mathematical ideal of perfect reconstruction.

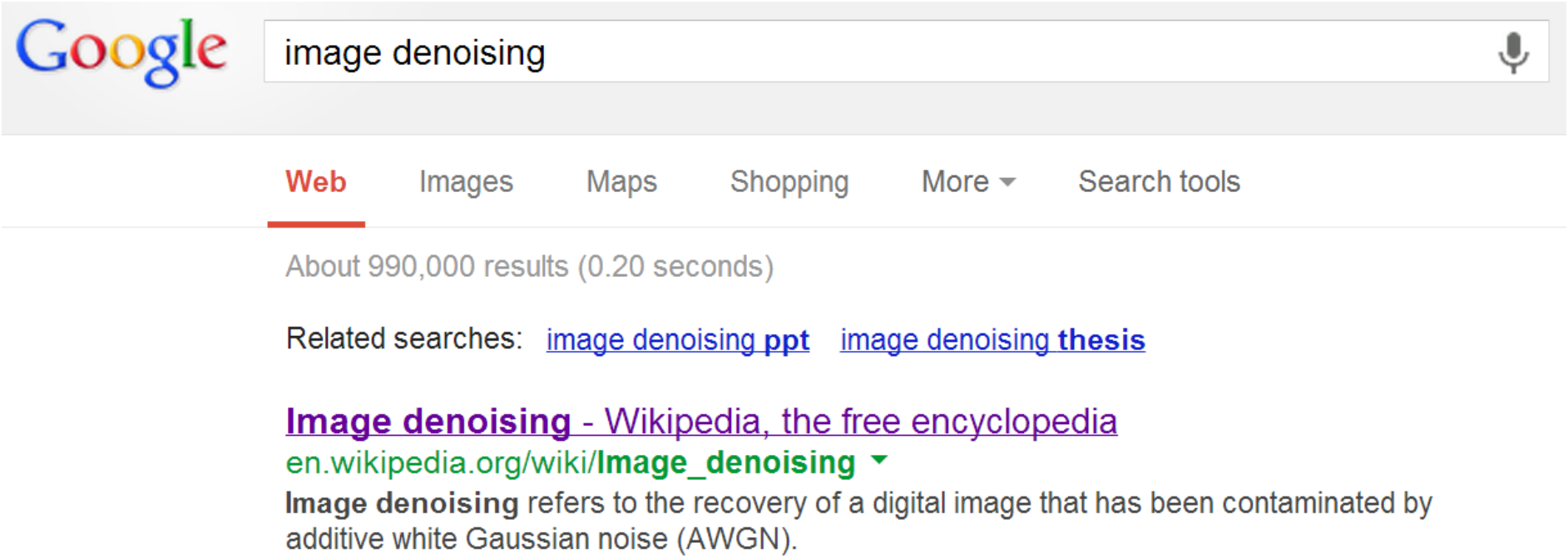
A Google User's Impression

- Viewed by many people ...



A Google User's Impression

- Searching for **image denoising** on Google ...



A Google User's Impression

- An article with limited information ...

Image denoising

From Wikipedia, the free encyclopedia

It has been suggested that this article be merged with *Noise reduction*. (Discuss) *Proposed since June 2010.*

This article **may be too technical for most readers to understand**. Please help improve this article to make it understandable to non-experts, without removing the technical details. The talk page may contain suggestions. *(July 2009)*

This article **does not cite any references or sources**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(April 2009)*

See also: Noise reduction#In images

Image denoising refers to the recovery of a digital image that has been contaminated by additive white Gaussian noise (AWGN).

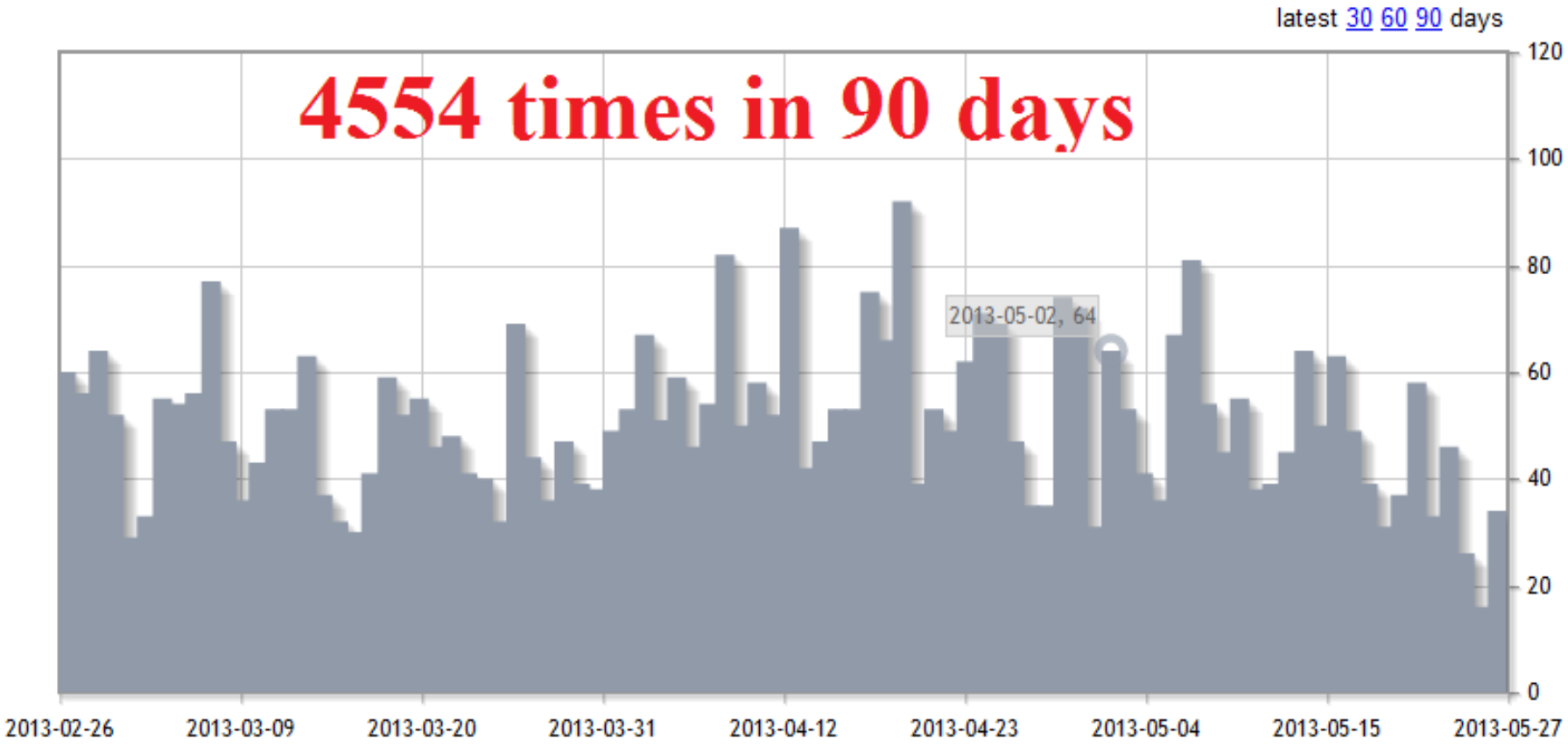
Contents [hide]

1 History
2 Technical description
3 Competing approaches
4 Applications

A Google User's Impression

- Viewed by some people ...

image denoising has been viewed 4554 times in the last 90 days.



Wikipedia and Signal Processing

- Wikipedia
  - A widely-used resource
  - Freelance editing model: anyone can edit

Wikipedia and Signal Processing

- Wikipedia
  - A widely-used resource
  - Freelance editing model: anyone can edit
- Signal Processing (SP) articles on Wikipedia
  - $>$1000 Articles, still growing
  - Grouped by subcategories
  - Need to monitor their quality!

Outline

- Ranking article importance

Outline

- Ranking article importance
- Assessing article quality

Outline

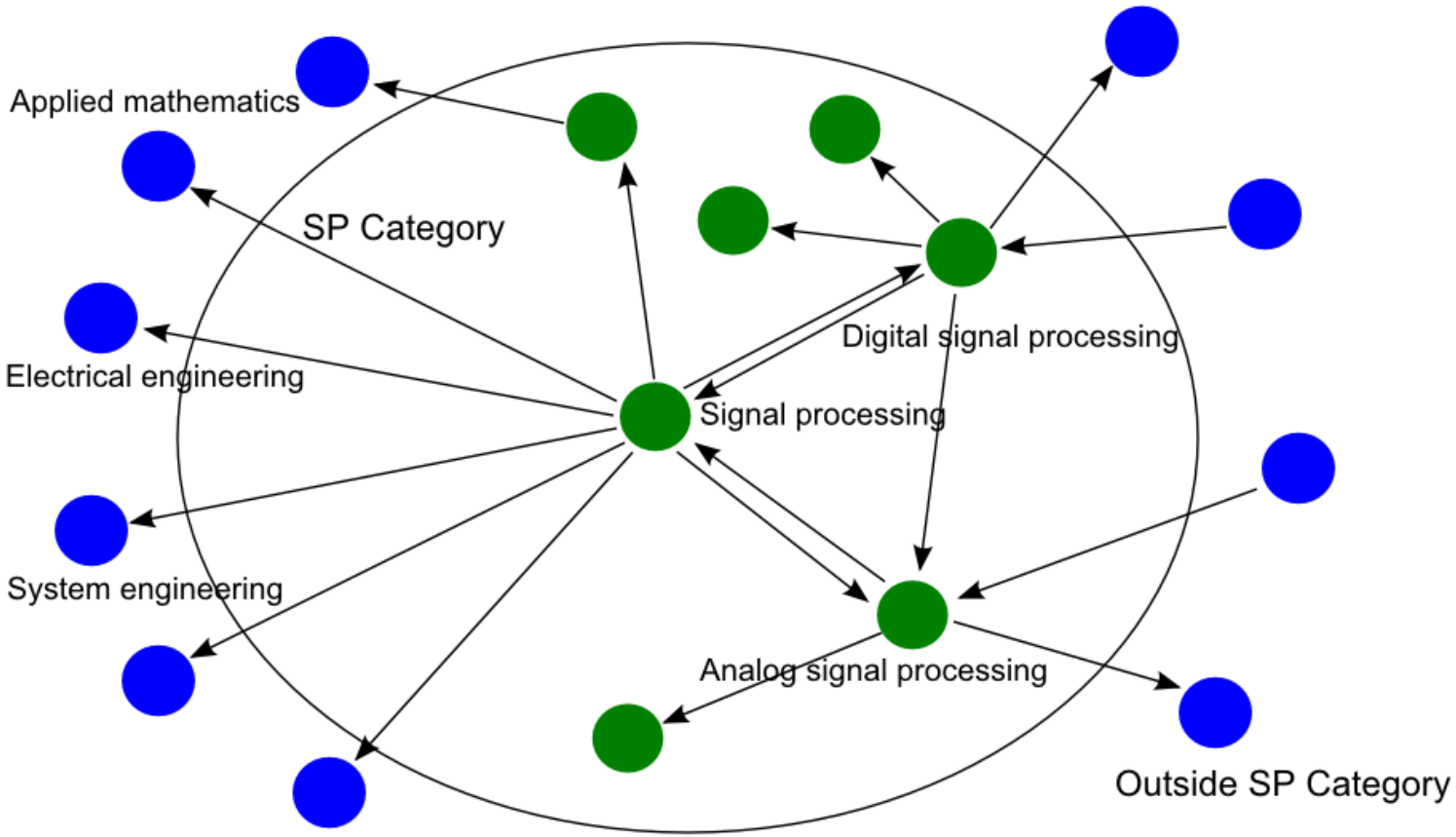
- Ranking article importance
- Assessing article quality
- Generating an improvement list

Outline

- Ranking article importance
- Assessing article quality
- Generating an improvement list
- Conclusions & future work

Importance Ranking: PageRank and HITS

How to rank SP articles on Wikipedia ...



Importance Ranking: PageRank and HITS

- PageRank [Brin98]
  - Rank the probability of visiting an article
  - A random walk model
  - An eigenvalue problem: find the eigenvector with eigenvalue 1 for a stochastic matrix

Importance Ranking: PageRank and HITS

- PageRank [Brin98]
  - Rank the probability of visiting an article
  - A random walk model
  - An eigenvalue problem: find the eigenvector with eigenvalue 1 for a stochastic matrix
- HITS [Kleinberg99]
  - Rank the authority of an article
  - Two scores
    - Authority: summation of hubness of point-to neighbors
    - Hubness: summation of authority of point-by neighbors
  - Iterative computation

Top-15 Articles by PageRank

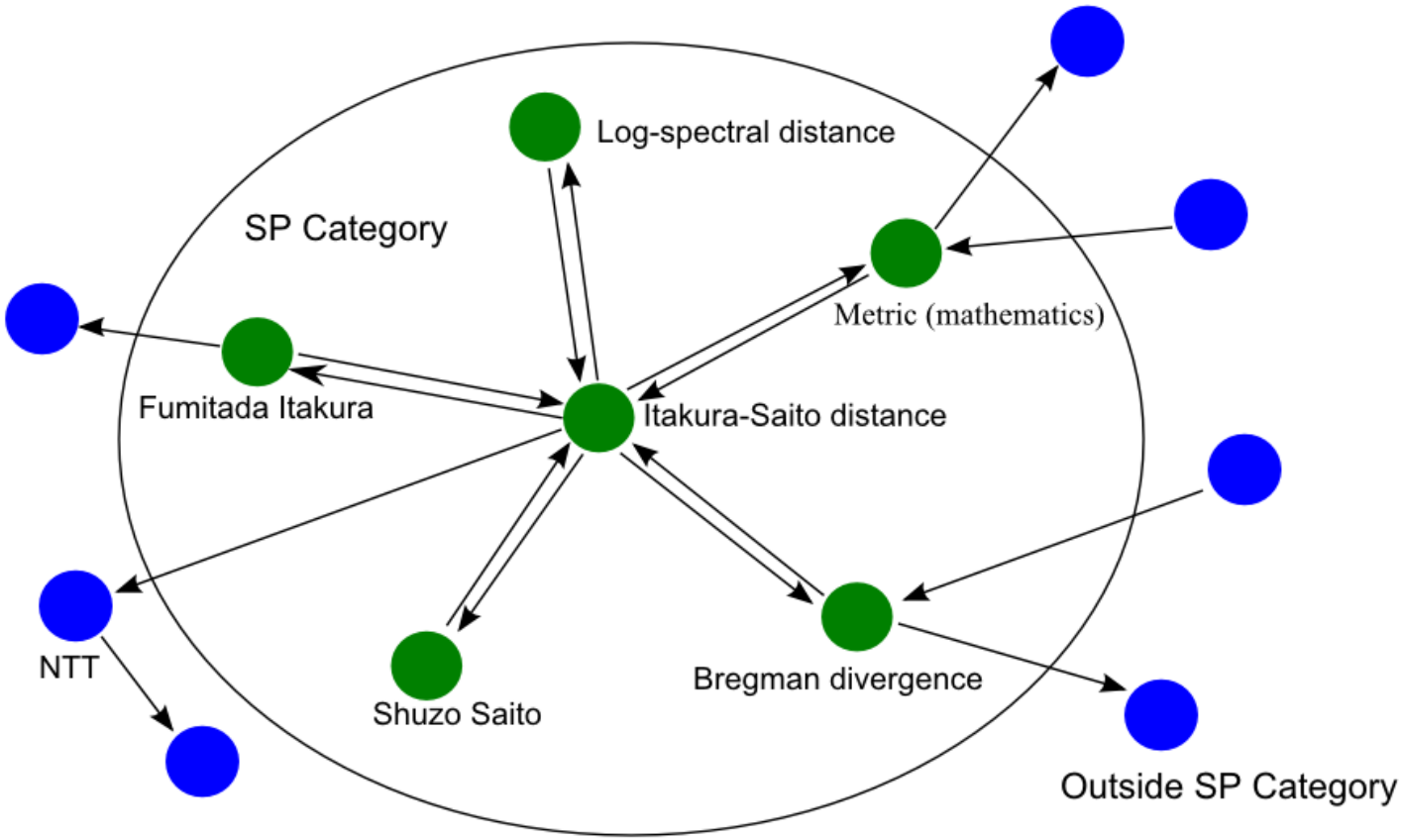
Ranking	Article
1	Kalman filter
2	Signal-to-noise ratio
3	Bilinear time–frequency distribution
4	Signal processing
5	Itakura–Saito distance
6	Ridge detection
7	Short-time Fourier transform
8	Thunder
9	Nyquist–Shannon sampling theorem
10	A-weighting
11	Image processing
12	Nyquist frequency
13	Hilbert transform
14	Wigner distribution function
15	Gaussian noise

Top-15 Articles by HITS

Ranking	Article
1	Dirac delta function
2	Dirac comb
3	Nyquist–Shannon sampling theorem
4	Whittaker–Shannon interpolation formula
5	Nyquist frequency
6	Fourier analysis
7	Discrete Fourier transform
8	Digital signal processing
9	Fast Fourier transform
10	LTI system theory
11	Kalman filter
12	Nyquist rate
13	Short-time Fourier transform
14	Discrete-time Fourier transform
15	Wiener filter

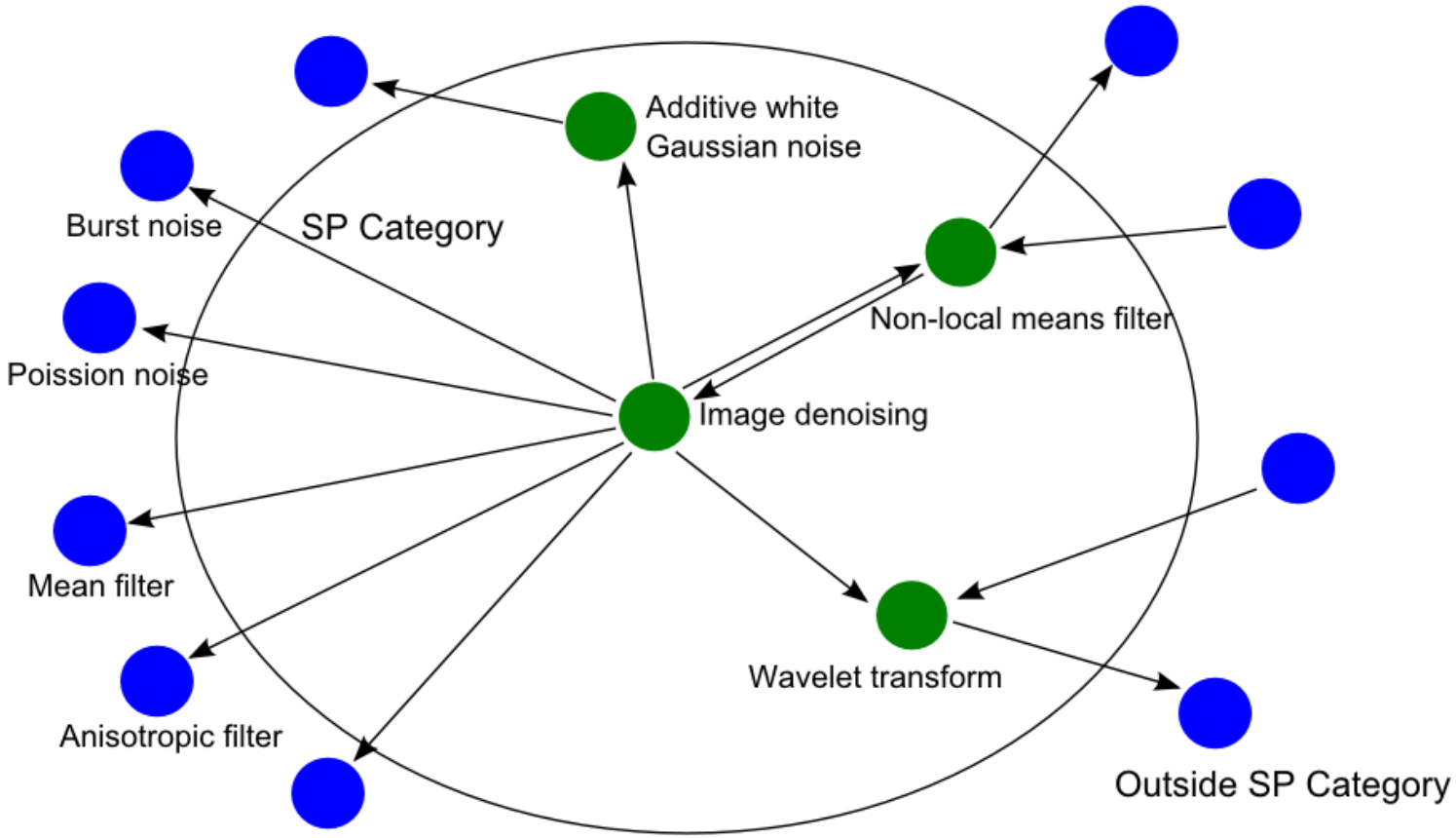
Island Structure: The Case of *Itakura–Saito Distance*

- Island structure is favored by PageRank



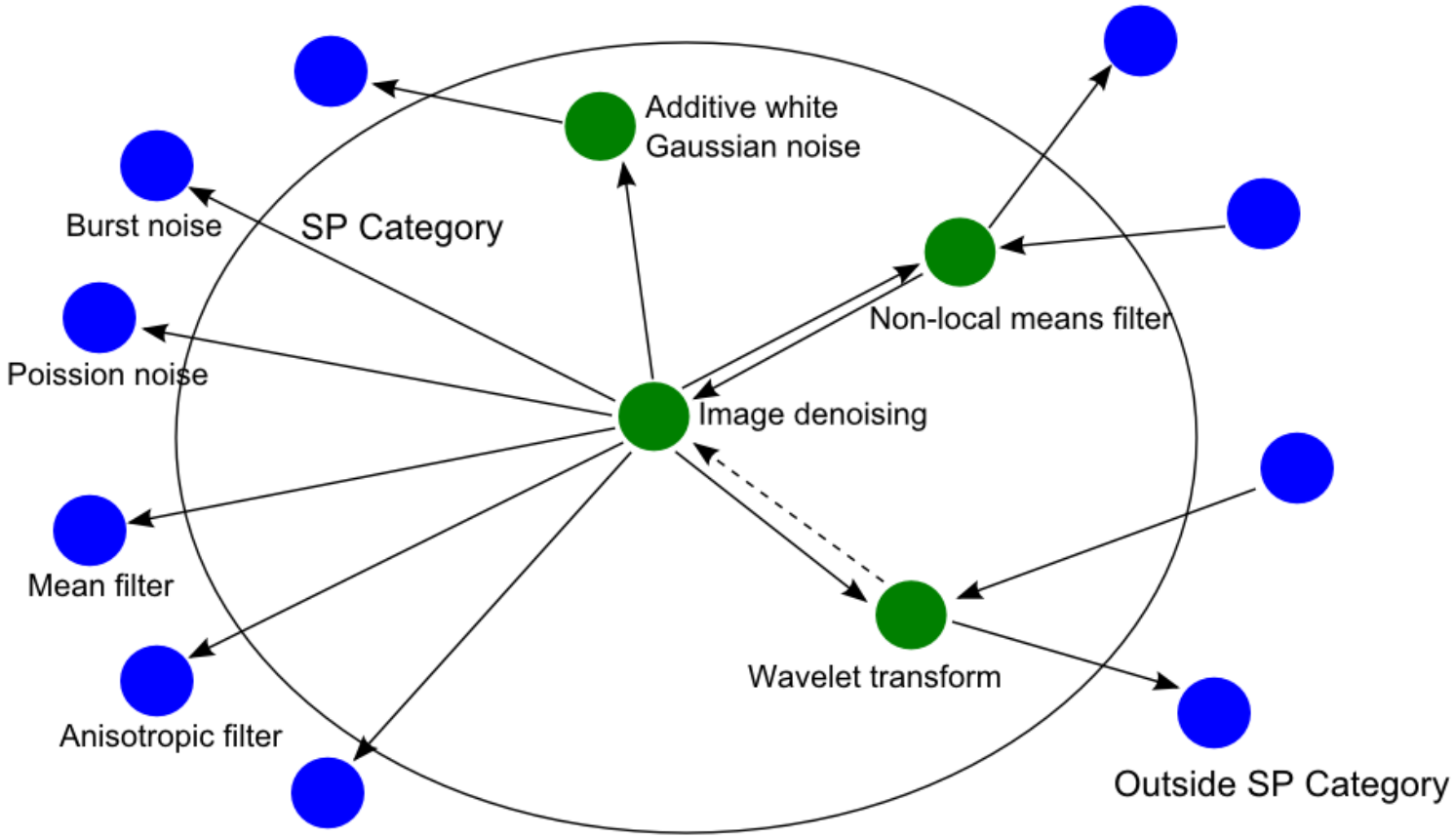
Where Is *Image Denoising*?

- Important but under-ranked



Where Is *Image Denoising*?

- Visibility can be improved by adding links



Importance Ranking via Crowdsourcing

- Contributed by 19/50 researchers from EPFL and elsewhere

Ranking	Article
1	Convolution
2	Fast Fourier transform
3	Nyquist-Shannon sampling theorem
4	Sampling (signal processing)
5	Filter (signal processing)
6	Fourier analysis
7	Kalman filter
8	Cross-correlation
9	Wavelet transform
10	Impulse response
11	Kalman filter
12	Discrete Fourier transform

Information Quality Analysis

- Heuristics-based metrics [Stvilia07]
  - Reputation
  - Completeness

Metric = Σ (Parameter $\cdot$ Weight)		
Metric	Parameter	Weight
Reputation	# editors	0.2
	# edits	0.2
	# articles connected through common editors	0.1
	# reverts	0.3
	# external links	0.2
	# registered user edits	0.1
	# anonymous user edits	0.2
Completeness	# internal links	0.4
	article length	0.6

Top-15 Articles by Reputation

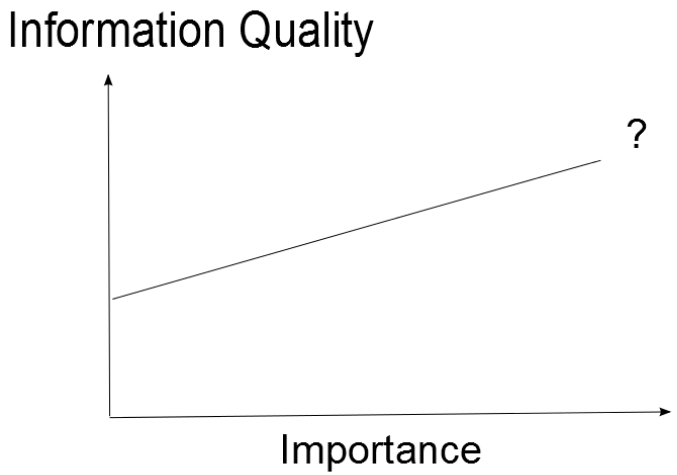
Ranking	Article
1	Analog-to-digital converter
2	Charge-coupled device
3	Convolution
4	Noise
5	Microelectromechanical systems
6	Sensor
7	Digital signal processing
8	Discrete Fourier transform
9	Pixel
10	Computer vision
11	Relay
12	White noise
13	Doppler effect
14	Dirac delta function
15	Potentiometer

Top-15 Articles by Completeness

Ranking	Article
1	Geophysical MASINT
2	Dirac delta function
3	Kalman filter
4	Avizo (software)
5	Noise in music
6	Allan variance
7	Mathematics of radio engineering
8	Discrete Fourier transform
9	Mechanical filter
10	JPEG 2000
11	Ordinary least squares
12	Color vision
13	Maximum likelihood
14	Hilbert transform
15	Nyquist–Shannon sampling theorem

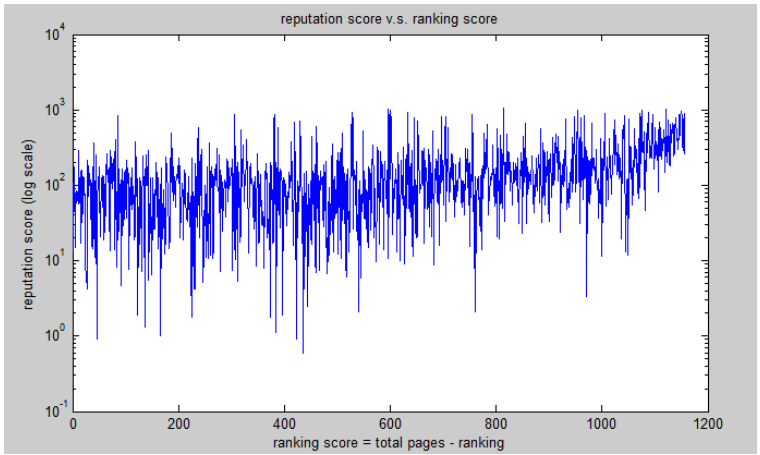
Information Quality v.s. Importance

- Scores
  - Importance score = (total articles - HITS ranking)
  - Information quality score = reputation/completeness scores
- Proportional?

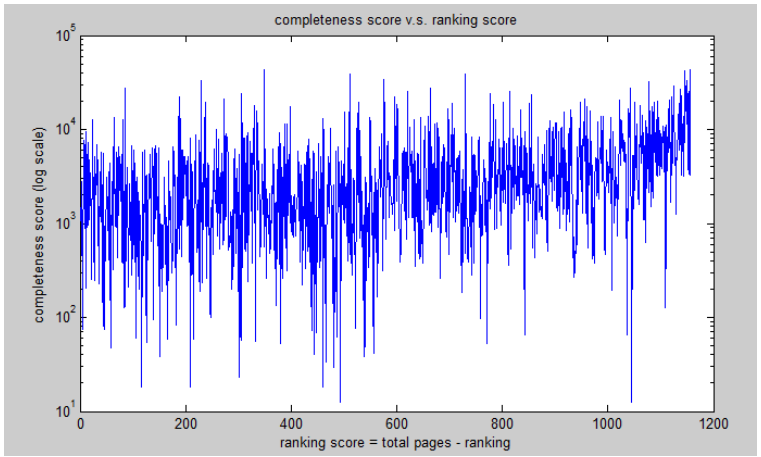


Information Quality v.s. Importance

- Strong fluctuations



(c) Reputation v.s. Importance



(d) Completeness v.s. Importance

Generating an Improvement List

- Articles to be improved
  - High ranking difference between importance and information quality
  - High importance ranking (high HITS ranking)
  - Still incomplete (low completeness score)

Generating an Improvement List

- Articles to be improved
  - High ranking difference between importance and information quality
  - High importance ranking (high HITS ranking)
  - Still incomplete (low completeness score)

- Need For Improvement (NFI) score

$$\text{NFI score} = \Gamma \cdot \theta(d) \cdot \delta(c)$$

where

$\Gamma = (\text{total articles} - \text{HITS ranking})$

$d = \text{difference score}, c = \text{completeness score}$

$$\theta(d) = \begin{cases} d & : d > threshold_{difference} \\ 0 & : \text{otherwise} \end{cases}$$

$$\delta(c) = \begin{cases} c & : c < threshold_{completeness} \\ 0 & : \text{otherwise} \end{cases}$$

Top-15 Articles on the Improvement List

Ranking	Article
1	Noise reduction
2	Continuous wavelets
3	Gabor limit
4	Gaussian noise
5	Modified Morlet wavelet
6	Noiselet
7	Spectral density estimation
8	Noise pollution
9	Noise spectral density
10	Periodic summation
11	Coherent sampling
12	N-jet
13	Bispectrum
14	Digital audio
15	Effective input noise temperature

$(threshold_{difference},\ threshold_{completeness}) = (50,\ 600)$

Top-15 Articles on the Improvement List

Noise reduction

From Wikipedia, the free encyclopedia

For the reduction of a sound's volume, see soundproofing. For the noise reduction of machinery and products, see noise control.

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(March 2010)*

Noise reduction is the process of removing noise from a signal.

All recording devices, both analogue or digital, have traits which make them susceptible to noise. Noise can be random or white noise with no coher coherent noise introduced by the device's mechanism or processing algorithms.

In electronic recording devices, a major form of noise is *hiss* caused by random electrons that, heavily influenced by heat, stray from their designatec These stray electrons influence the voltage of the output signal and thus create detectable noise.

In the case of photographic film and magnetic tape, noise (both visible and audible) is introduced due to the grain structure of the medium. In photogr film, the size of the grains in the film determines the film's sensitivity, more sensitive film having larger sized grains. In magnetic tape, the larger the g the magnetic particles (usually ferric oxide or magnetite), the more prone the medium is to noise.

To compensate for this, larger areas of film or magnetic tape may be used to lower the noise to an acceptable level.

Contents [hide]

Top-15 Articles on the Improvement List

Spectral density estimation

From Wikipedia, the free encyclopedia

It has been suggested that this article be merged into *Frequency domain*. (Discuss) *Proposed since August 2012.*

In statistical signal processing, the goal of **spectral density estimation** is to estimate the spectral density (also known as the power spectrum) of a signal from a sequence of time samples of the signal. Intuitively speaking, the spectral density characterizes the frequency content of the signal. The purpose of estimating the spectral density is to detect any periodicities in the data, by observing peaks at the frequencies corresponding to these periodicities.

SDE should be distinguished from the field of frequency estimation, which assumes a limited (usually small) number of generating frequencies plus and seeks to find their frequencies. SDE makes no assumption on the number of components and seeks to estimate the whole generating spectrum

Techniques [edit]

Techniques for spectrum estimation can generally be divided into *parametric* and *non-parametric* methods. The parametric approaches assume th underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an regressive or moving average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic proce contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any pa structure.

Conclusions

- Importance and quality of articles are mismatched

	High Importance	Low Importance
Good quality	Nyquist–Shannon sampling theorem	
Bad quality		

Conclusions

- Importance and quality of articles are mismatched

	High Importance	Low Importance
Good quality	Nyquist–Shannon sampling theorem	
Bad quality	Gaussian noise	

Conclusions

- Importance and quality of articles are mismatched

	High Importance	Low Importance
Good quality	Nyquist–Shannon sampling theorem	Avizo (software)
Bad quality	Gaussian noise	

Conclusions

- Importance and quality of articles are mismatched

	High Importance	Low Importance
Good quality	Nyquist–Shannon sampling theorem	Avizo (software)
Bad quality	Gaussian noise	AutoCollage 2008

Conclusions

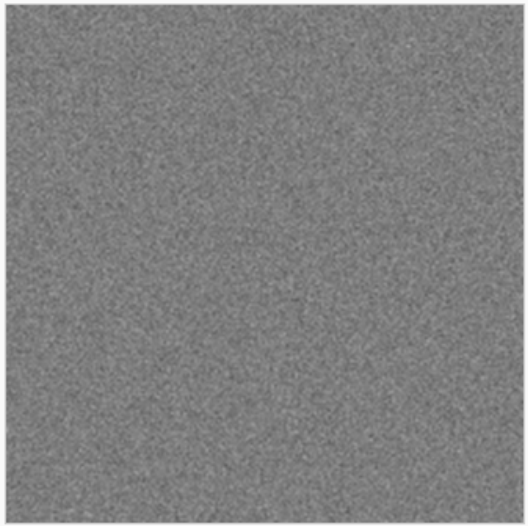
- Some important articles are highlighted for improvement

Gaussian noise

From Wikipedia, the free encyclopedia

This article **relies largely or entirely upon a single source**. Relevant discussion may be found on the talk page. Please help improve this article by introducing citations to additional sources. *(October 2012)*

Gaussian noise is statistical noise that has its probability density function equal to that of the normal distribution, which is also known as the Gaussian distribution.[1] In other words, the values that the noise can take on are Gaussian-distributed. A special case is **white Gaussian noise**, in which the values at any pair of times are identically distributed and statistically independent (and hence uncorrelated). In applications, Gaussian noise is most commonly used as additive white noise to yield additive white Gaussian noise.



A grayscale image created from gaussian noise.

See also [edit]

- Gaussian process

References [edit]

1. ^ Barry Truax, ed. (1999). "Handbook for Acoustic Ecology" (Second ed.). Cambridge Street Publishing. Retrieved 2012-08-05.

Conclusions

- Visibility of articles could be improved by adding links

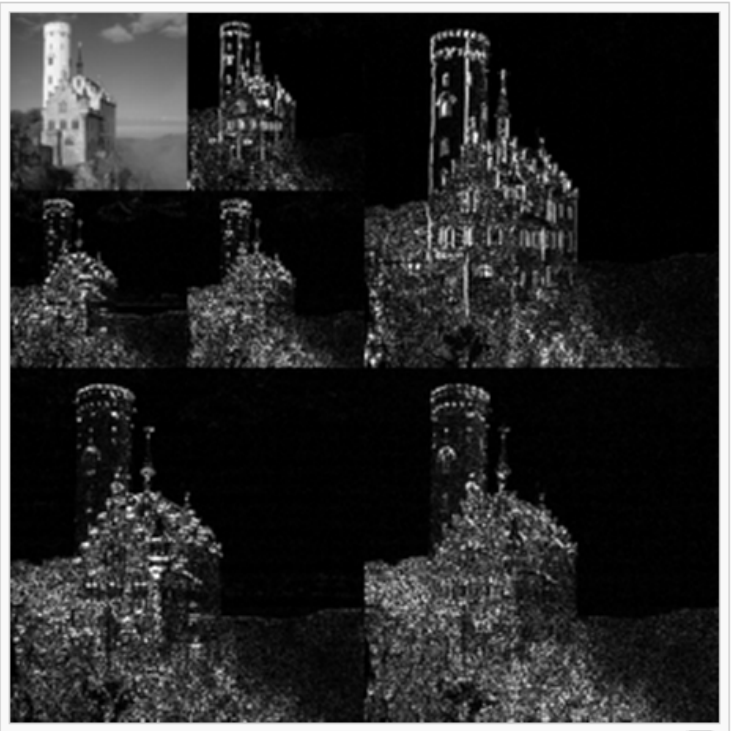
Wavelet transform

From Wikipedia, the free encyclopedia

In mathematics, a **wavelet series** is a representation of a square-integrable (real- or complex-valued) function by a certain orthonormal series generated by a wavelet. Nowadays, wavelet transformation is one of the most popular candidates of the time-frequency-transformations. This article provides a formal, mathematical definition of an **orthonormal wavelet** and of the **integral wavelet transform**.

Contents [hide]

Image Denoising



Conclusions

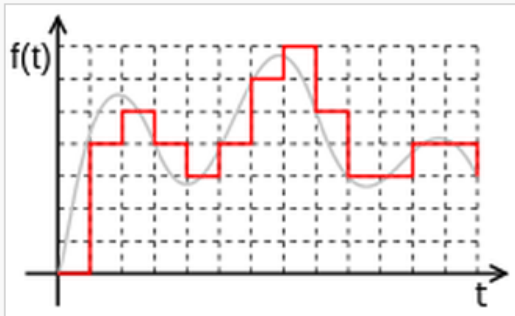
- Audio/speech articles could benefit from further improvement

Digital audio

From Wikipedia, the free encyclopedia

"Digital music" redirects here. For a kind of modern music composed by digital means, see electronic music.

Digital audio refers to technology that records, stores, and reproduces sound by encoding an audio signal in digital form instead of analog form. Sound is passed through an analog-to-digital converter (ADC), and pulse-code modulation is typically used to encode it as a digital signal. A digital-to-analog converter performs the reverse process, and converts the digital signal back into an audible sound. Digital audio systems may include compression, storage, processing and transmission components. Conversion to a digital format allows convenient manipulation, storage, transmission and retrieval of an audio signal.



A sound wave, in gray, represented digitally, in red (after quantization and zero-order hold).

Contents [hide]

Future Work

- Multiple articles dealing with the same topic could be merged

Continuous wavelet transform

From Wikipedia, the free encyclopedia

This article **needs additional citations** article by adding citations to reliable sou and removed. *(June 2012)*

A **continuous wavelet transform** (CWT) is used to divide a continuous-time function into wavelets. Unlike Fourier transform, the continuous wavelet transform possesses the ability to construct a time-frequency representation of a signal that offers very good time and frequency localization. In mathematics, the continuous wavelet transform of a continuous, square-integrable function $x(t)$ at a scale $a > 0$ and translational value $b \in \mathbb{R}$ is expressed by the following integral

$$X_w(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt$$

Continuous wavelet

From Wikipedia, the free encyclopedia
(Redirected from Continuous wavelets)

This article **does not cite any** article by adding citations to reli and removed. *(December 2009)*

In numerical analysis, **continuous wavelets** are functions defined as analytical expressions, as functions either of tin both wavelet decomposition and composition transforms.

The following continuous wavelets have been invented for

- Morlet wavelet
- Modified Morlet wavelet
- Mexican hat wavelet
- Complex Mexican hat wavelet
- Shannon wavelet
- Difference of Gaussians

Future Work

- Exploring the interaction with other categories (e.g. mathematics)

Signal processing

From Wikipedia, the free encyclopedia

This article **needs additio**
article by adding citations t
and removed. *(December 20*

"Signal theory" redirects here. It is not to be coni

Signal processing is an area of systems engineering, electrical engineering and applied mathematics that deals with operations on or analysis of signals, or measurements of time-varying or spatially varying physical quantities. Signals of interest can include sound, electromagnetic radiation, images, and sensor data, for example biological data such as electrocardiograms, control system signals, telecommunication transmission signals, and many others.

Functional analysis

From Wikipedia, the free encyclopedia

For functional analysis as used in psychology, see functional analy

Functional analysis is a branch of mathematical analysis, the core of wh with some kind of limit-related structure (e.g. inner product, norm, topolog spaces and respecting these structures in a suitable sense. The historical of functions and the formulation of properties of transformations of functic defining continuous, unitary etc. operators between function spaces. This the study of differential and integral equations.

The usage of the word *functional* goes back to the calculus of variations, the name was first used in Hadamard's 1910 book on that subject. Howev been introduced in 1887 by the Italian mathematician and physicist Vito V continued by students of Hadamard, in particular Fréchet and Lévy. Hadar functional analysis further developed by Riesz and the group of Polish ma

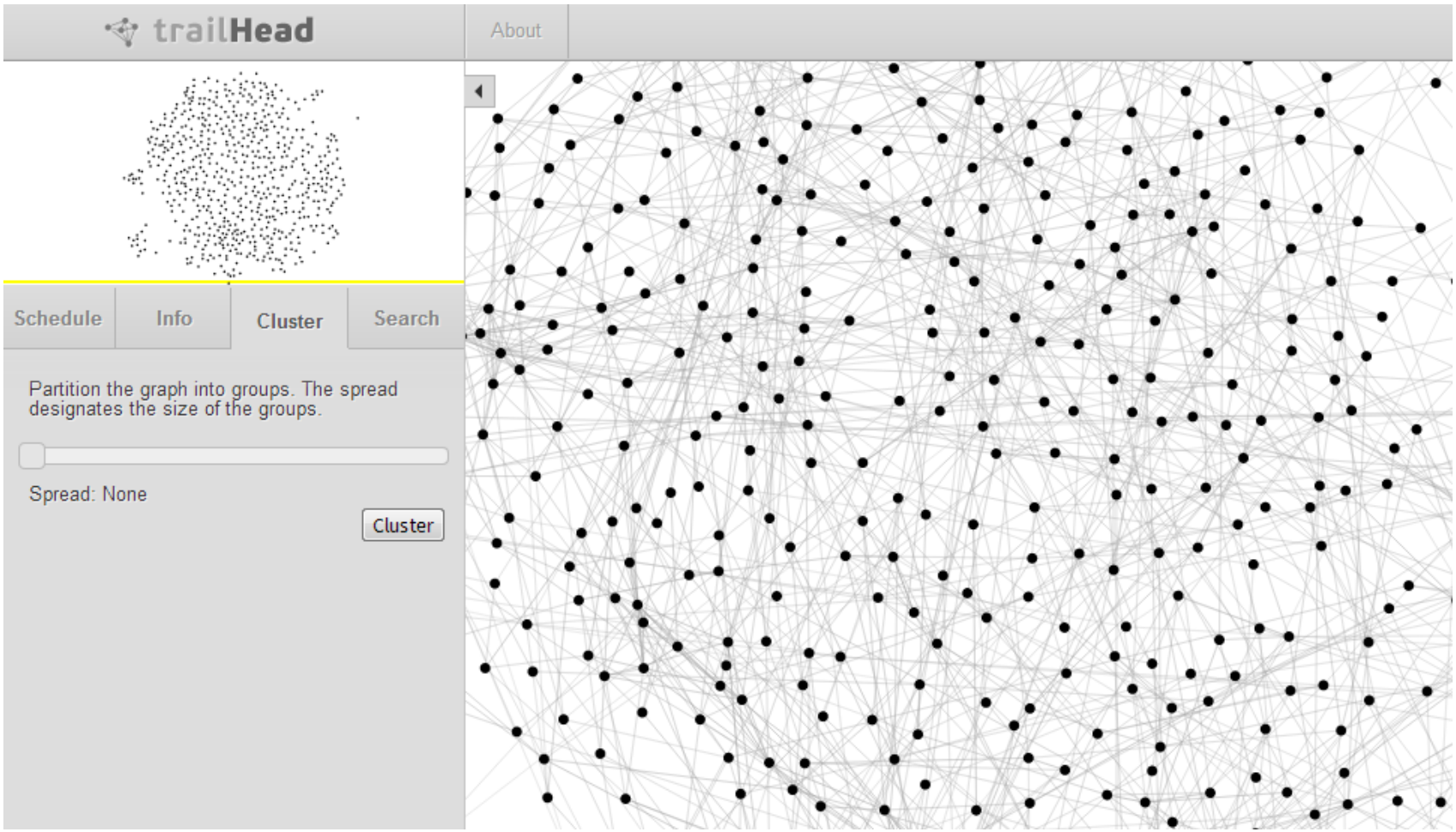
In modern introductory texts to functional analysis, the subject is seen as t in particular infinite dimensional spaces. In contrast, linear algebra deals n use topology. An important part of functional analysis is the extension of th infinite dimensional spaces, also known as **infinite dimensional analysis**

Future Work

- Crowdsourcing of article quality

Future Work

- Realtime monitoring of SP articles (trailHead as a starting point)



Thank you, questions please.

{tao.lee, jay.unnikrishnan}@epfl.ch

Software, dataset, results are available at
http://lcav.epfl.ch/page-87349-en.html

AUDIOVISUAL COMMUNICATIONS LCAV

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE