



TECH breakFast

by asseco

„Świat jest mały”

czyli wstęp do podstaw elementów
analizy sieci złożonych

Włodzimierz Kozłowski, Starszy Projektant

Wlodzimierz.Kozlowski@assecO.pl

wlodzimierz@kozlowscy.org

<https://www.linkedin.com/in/wlodzimierz-kozlowski>

- Co to są sieci małych światów i czym jest **liczba Bacona** (albo Erdősa)?
- Jakie są własności tzw. **sieci rzeczywistych** i jak w kilku liniach w **języku R** „zbudować” sobie Facebooka?
- Czym jest **rozkład potęgowy** i jak wykorzystać tę wiedzę do walki z epidemią albo przewidzieć skutki upadłości banku?
- Kto jest najważniejszy i jak to sprawdzić w **Sparku**?

....Pytania?

Zadawaj je, komentuj lub polub
poprzez sli.do:

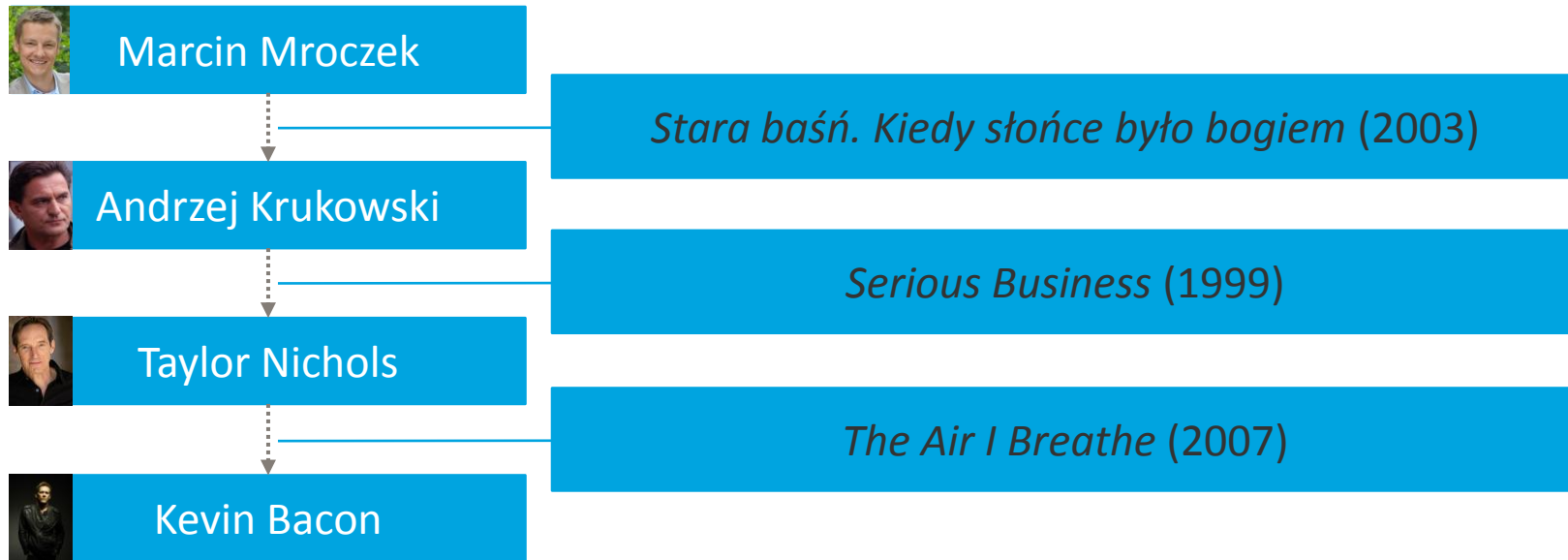
TechBreakfast.Asseco.com/zapytaj

Eksperyment Milgrama

*Eksperyment Milgrama dowiódł prawdziwości obiegowego porzekadła, że **świat jest mały**. Mimo że sieć społeczna liczy kilka miliardów ludzi, to średnia droga między dowolną parą węzłów w takiej sieci wynosi **około sześciu**. Ponad dwadzieścia lat po eksperymencie Milgrama ukute zostało nawet sformułowanie „sześć uścisków dłoni” albo „sześć stopni separacji”.*

źródło: Agata Fronczak, Piotr Fronczak. *Świat sieci złożonych. Od fizyki do Internetu*. Wydawnictwo Naukowe PWN, Warszawa 2009

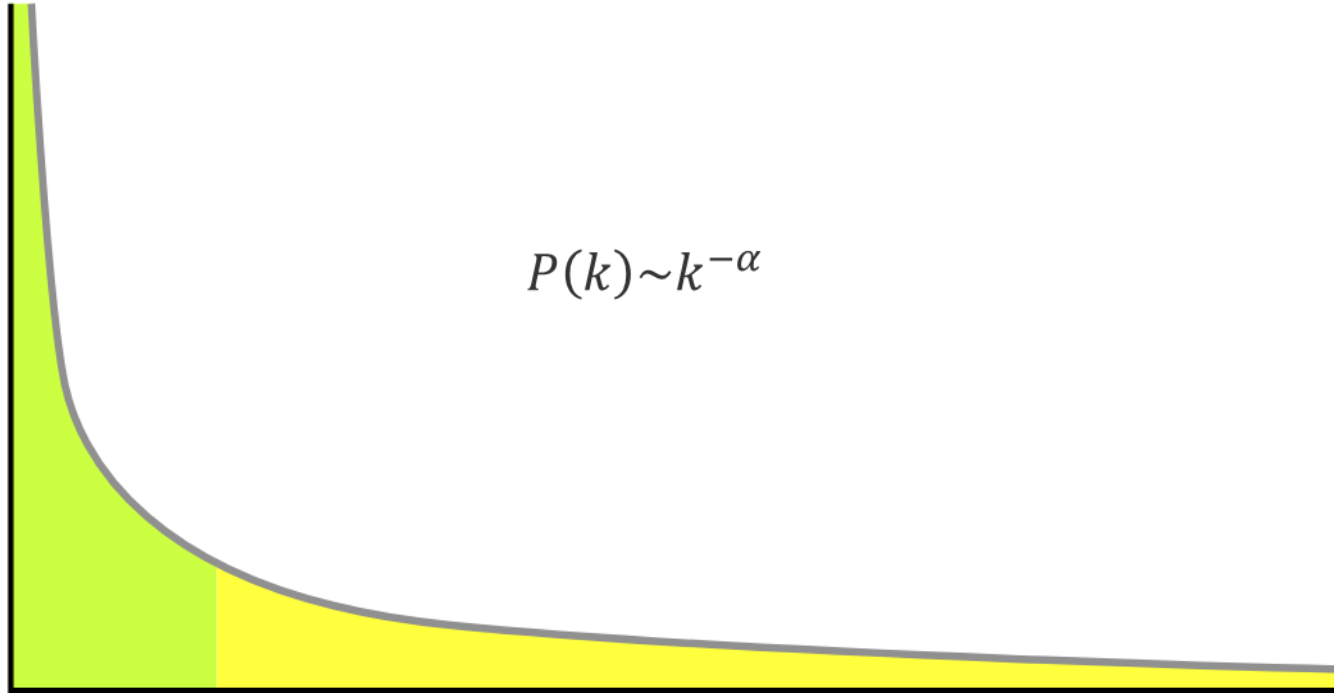
Liczba Bacona



źródło: Agata Fronczak, Piotr Fronczak. *Świat sieci złożonych. Od fizyki do Internetu*. Wydawnictwo Naukowe PWN, Warszawa 2009

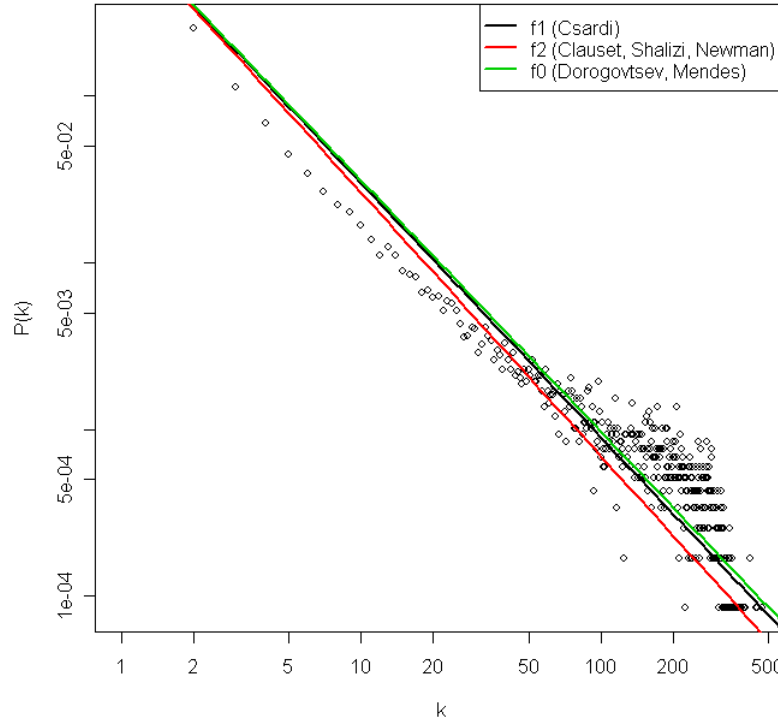
foto: <https://www.imdb.com>, <https://fdb.pl>

Rozkład potęgowy



rys.: https://en.wikipedia.org/wiki/Power_law

Rozkład potęgowy



źródło: Włodzimierz Kozłowski, *Wybrane elementy analizy sieci złożonych na przykładzie...*, Uniwersytet im. Adama Mickiewicza w Poznaniu, wrzesień 2007

Własności sieci rzeczywistych

- „małe światy”
- **potęgowy rozkład stopni wierzchołków**
- silnie zgronowane
- rzadkie
- ...

Przykłady sieci rzeczywistych

- aktorzy i cytowania naukowe,
- sieci społeczne i Internet,
- połączenia lotnicze,
- kontakty seksualne,
- transakcje międzybankowe,
- pliki w repozytorium,
- ...

i zastosowania?

igraph



igraph – The network analysis package

*igraph is a collection of network analysis tools with the emphasis on **efficiency**, **portability** and ease of use.*

*igraph is **open source** and free.*

*igraph can be programmed in **R**, **Python** and **C/C++**.*

Kurs R w jedną minutę...

```
> imie <- "Włodzimierz"  
> "Kozłowski" -> nazwisko  
> paste(imie, nazwisko)  
[1] "Włodzimierz Kozłowski"
```



więcej: <https://cran.r-project.org/doc/manuals/r-release/R-intro.html>

Kurs R w jedną minutę...



```
> hello <- function(imie, n = 1) {  
+   t <- paste("Wit", strrep("a", n), "j ", imie, "!", sep = "")  
+   print(t)  
+ }  
> hello("Paulina")  
[1] "Witaj Paulina!"  
> hello("Paulina", 3)  
[1] "Witaaaj Paulina!,,  
  
> type = c("gnp", "gnm")
```

więcej: <https://cran.r-project.org/doc/manuals/r-release/R-intro.html>

erdos.renyi.game

```
> ?erdos.renyi.game
```

```
# Generate random graphs according to the Erdos-Renyi model
```

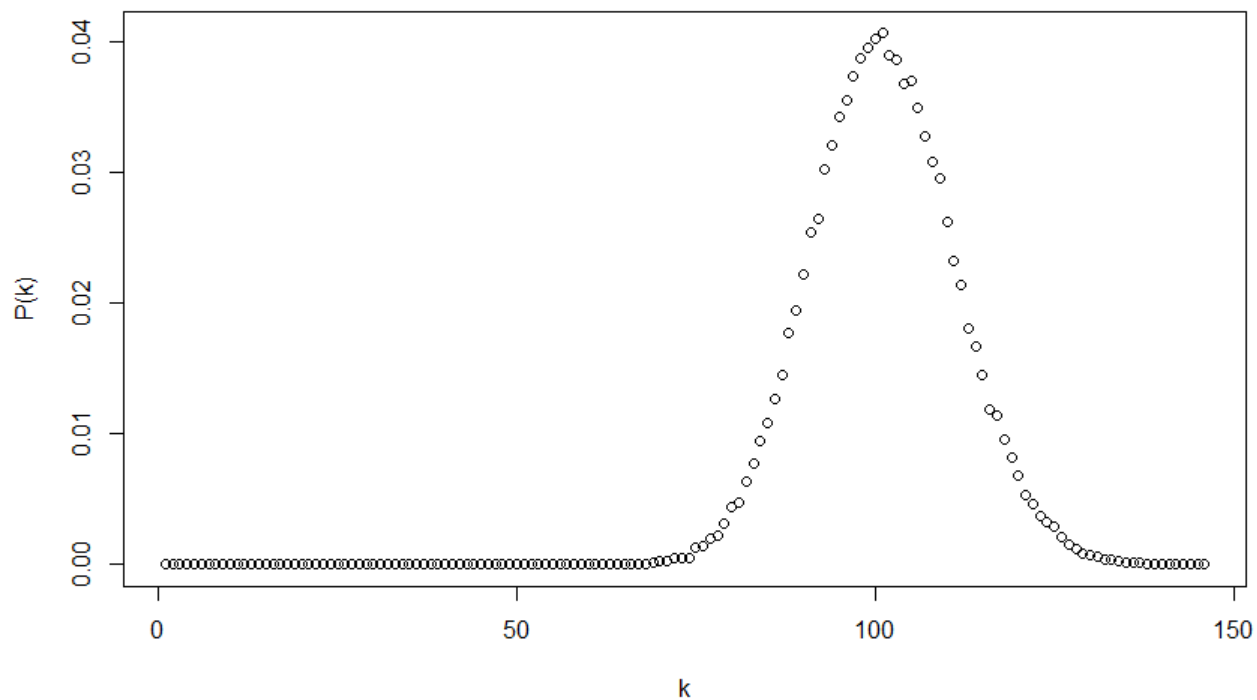
```
# This model is very simple, every possible edge is created with the same  
constant probability.
```

```
# random.graph.game are deprecated, and sample_gnp and sample_gnm should be  
used instead
```

erdos.renyi.game

```
# utworzenie sieci
> ER <- erdos.renyi.game(100000, 1/1000)
> average.path.length(ER)
[1] 2.902858
> diameter((ER))
[1] 4
# rozkład stopni wierzchołków i wykres
> ER$Pk <- degree.distribution(ER)
> plot(ER$Pk, xlab = "k", ylab = "P(k)")
```

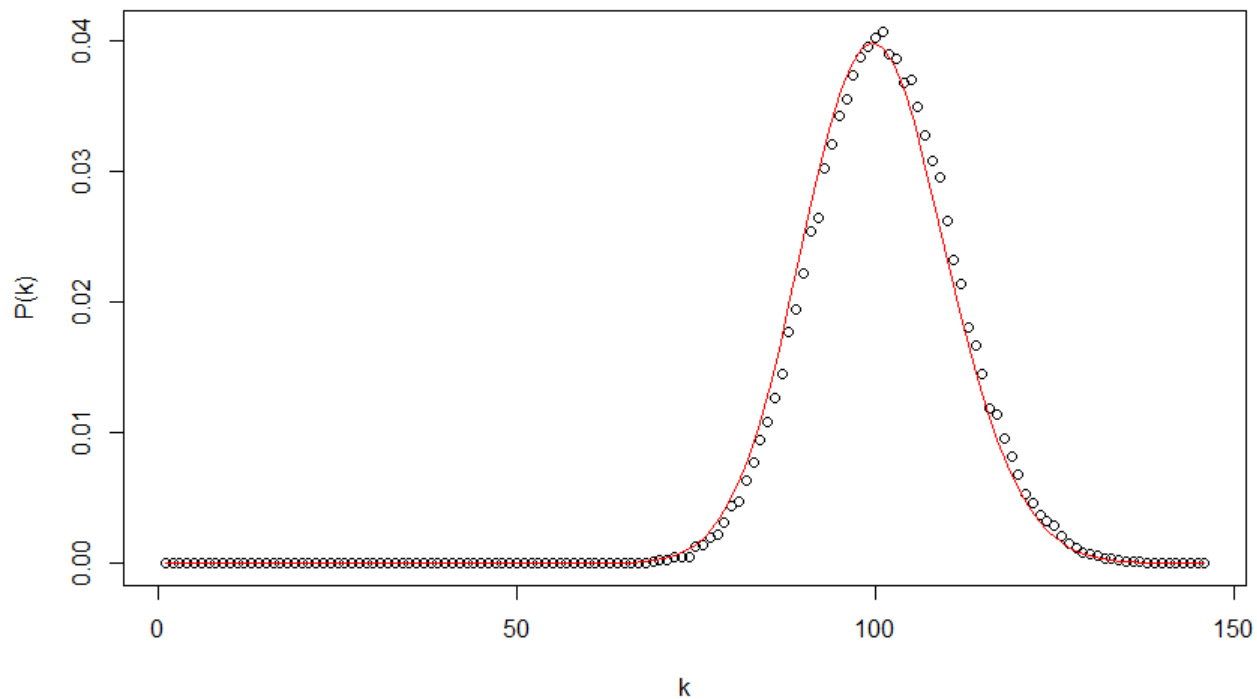
erdos.renyi.game



erdos.renyi.game

```
# stopnie wierzchołków  
> ER$k <- degree(ER)  
# rozkład Poissona  
> lambda <- mean(ER$k)  
> curve(((exp(-lambda) * lambda^(x))/factorial(x)), col = "red", add = TRUE)
```

erdos.renyi.game



barabasi.game

```
> ?barabasi.game
```

```
# Generate scale-free graphs according to the Barabasi-Albert model
```

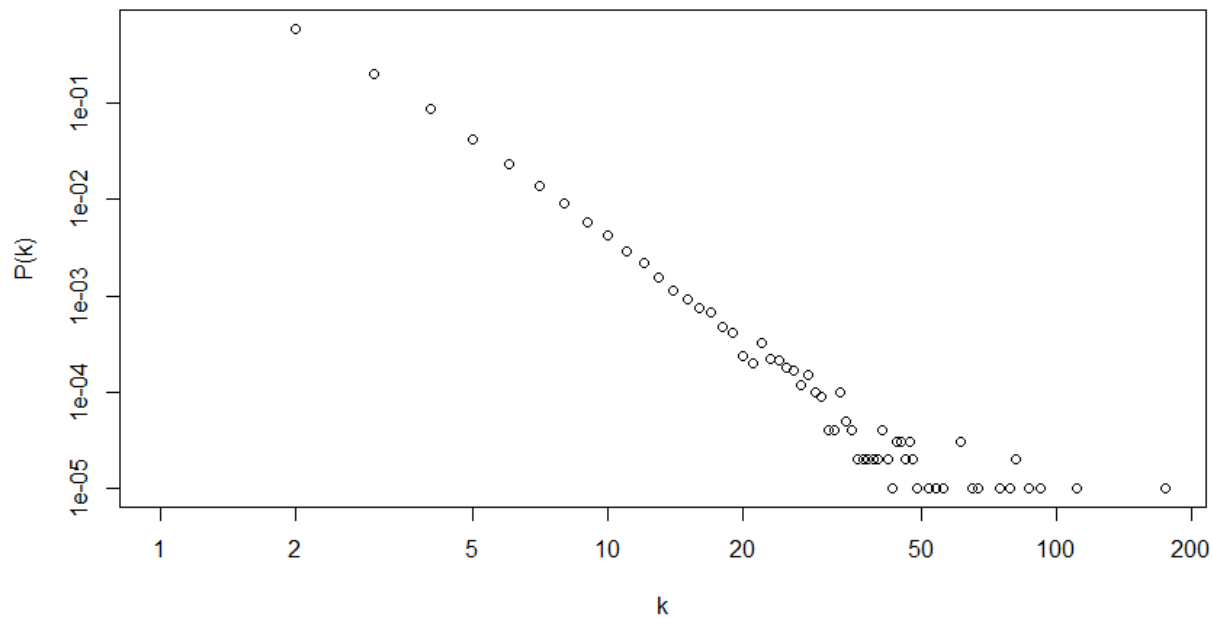
```
# The BA-model is a very simple stochastic algorithm for building a graph.
```

```
# sample_pa
```

barabasi.game

```
# utworzenie sieci
> BA <- barabasi.game(100000, directed = FALSE)
> average.path.length(BA)
[1] 13.70432
> diameter(BA)
[1] 35
# rozkład stopni wierzchołków i wykres
> BA$Pk <- degree.distribution(BA)
> plot(BA$Pk, xlab = "k", ylab = "P(k)", log = "xy")
```

barabasi.game



barabasi.game

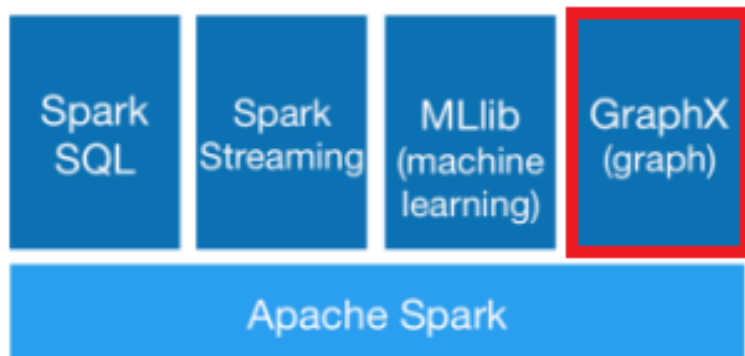
```
# określenie współczynnika alfa
> BA$k <- degree(BA)
> y <- power.law.fit(BA$k, implementation = "R.mle")
> y
(...)
Coefficients:
      alpha
2.187766

> write.graph(BA, "c:/temp/BA-100k", format = "edgelist")
```

Apache Spark GraphX



GraphX is Apache Spark's API for graphs and graph-parallel computation.



```
object Graph {
  def apply[VD, ED](
    vertices: RDD[(VertexId, VD)],
    edges: RDD[Edge[ED]],
    defaultVertexAttr: VD = null)
    : Graph[VD, ED]
  (...)
}
```

GraphFrames

GraphX is to RDDs as GraphFrames are to DataFrames.

GraphFrames also provide powerful tools for running queries and standard graph algorithms. With GraphFrames, you can easily search for patterns within graphs, find important vertices, and more.

źródło: <https://graphframes.github.io/index.html>

więcej: <https://databricks.com/blog/2016/03/03/introducing-graphframes.html>

GraphFrames

```
import org.graphframes.examples
val g: GraphFrame = examples.Graphs.friends // get example graph

// Search for pairs of vertices with edges in both directions between them.
val motifs: DataFrame = g.find("(a)-[e]->(b); (b)-[e2]->(a)")
motifs.show()

// More complex queries can be expressed by applying filters.
motifs.filter("b.age > 30").show()
```

PageRank

Kto jest najważniejszy?

Ważny nie jest ten, który ma dużo znajomych ale ten, który ma ważnych znajomych.

więcej: <https://en.wikipedia.org/wiki/PageRank>

PageRank

```
val customSchema = StructType(Array(
  StructField("src", IntegerType, true),
  StructField("dst", IntegerType, true)))
val df = sqlContext.read.format("com.databricks.spark.csv")
  .option("header", "false").option("delimiter", ",")
  .schema(customSchema)
  .load("BA-100k")

val n1 = edges.select("src").distinct()
val n2 = edges.select("dst").distinct()
val n = n1.unionAll(n2).withColumnRenamed("src", "name").distinct()
val nodes = n.withColumn("id", n("name"))

val g = GraphFrame(nodes, edges)
val pr = g.pageRank.resetProbability(0.15).maxIter(10).run()
pr.vertices.sort(desc("pagerank")).show()
```

źródło: <https://blog.cloudera.com/blog/2016/10/how-to-do-scalable-graph-analytics-with-apache-spark>

Literatura

Agata Fronczak, Piotr Fronczak. *Świat sieci złożonych. Od fizyki do Internetu*. Wydawnictwo Naukowe PWN, Warszawa 2009.

S.N. Dorogovtsev, J. F. F. Mendes. *Evolution of Networks. From Biological Nets to the Internet and WWW*. Oxford 2003.

git clone



<https://github.com/wkozi/swiat-jest-maly>

@author

„Programujący projektant” w Pionie Energetyki i Gazownictwa Asseco Poland S.A.

Od kilkunastu lat związany z największym systemem bilingowym w polskiej energetyce a przede wszystkim ludźmi, którzy go tworzą. Ostatnio oddelegowany także do realizacji projektu „Badania nad zastosowaniem technologii Internet-of-Things oraz opracowaniem modeli regulacji przy użyciu różnych urządzeń, w tym akumulatorów domowych i pojazdów elektrycznych...”.

Od jakiegoś czasu żywo zainteresowany technologiami „dużych danych”. Absolwent studiów podyplomowych „Big Data – przetwarzanie i analiza dużych zbiorów danych” i wkrótce uczestnik „Data Science – algorytmy, narzędzia i aplikacje dla problemów typu Big Data” na Wydziale Elektroniki i Technik Informacyjnych Politechniki Warszawskiej.

Prywatnie, oprócz standardowego zestawu: praca, żona (1) i dzieci (3), instruktor ZHR (zhr.pl) i „wędrownik” (wedrownicy.org), który zabiera plecak i samotnie wędruje zimą po Uralu albo rowerem po Górskim Karabachu.

Solutions
for demanding
business.

asreco

Zastrzeżenia prawne

Zawartość dostępna w prezentacji jest chroniona prawem autorskim i stanowi przedmiot własności. Teksty, grafika, fotografie, dźwięk, animacje i filmy, a także sposób ich rozmieszczenia w prezentacji podlegają ochronie na mocy Ustawy o prawach autorskich i prawach pokrewnych oraz innych przepisów z tym związanych. Jakiegokolwiek nieautoryzowane zastosowanie jakichkolwiek materiałów zawartych w prezentacji może stanowić naruszenie praw autorskich, znaków firmowych lub innych przepisów. Materiały dostępne w prezentacji nie mogą być modyfikowane, powielane, przedstawiane publicznie, wykonywane, rozprowadzane lub wykorzystywane w innych celach publicznych lub komercyjnych, chyba że Asseco Poland S.A. wydał na to wyraźną zgodę na piśmie. Kopiowanie w celach komercyjnych, rozpowszechnianie, modyfikacja lub przejmowanie zawartości niniejszej prezentacji przez osoby trzecie jest niedozwolone. W prezentacji mogą być prezentowane również materiały zawierające odesłania do ofert i usług podmiotów trzecich. Warunki korzystania z ofert i usług podmiotów trzecich są określone przez te podmioty. Asseco Poland S.A. nie ponosi żadnej odpowiedzialności za warunki i skutki korzystania z ofert i usług tychże podmiotów. Dane i informacje zawarte w prezentacji mają jedynie charakter ogólnoinformacyjny. Prezentacja przygotowana została w oparciu i przy użyciu produktów firmy Inscale.

Nazwa oraz logo Asseco Poland S.A. są zarejestrowanymi znakami towarowymi. Korzystanie z tych znaków wymaga wyraźnej zgody ze strony Asseco Poland S.A.